



```

124 LysLeuValProTrpMetIleLeuGlySerLeuLeuTyValSerMetI 140
152 AGCGTGCCTCCATGATGATCTGGGCTCTCGCATATGATCATCAT 201
140 eCyValPheHisSerLysTyValGlyPheMetValProTyPheLeu 157
202 TTGTGTTTCCATGCAAAATATGACAGGGTTATGCTCCATCTTCCT 251
157 rGlyPhePheSerGlnAsnAlaThrIleGlnLysGluAspThrLeu 173
252 GGAATTTTTCCTCCCAAAATGCCAATTCAAAAGAAATACACTGGCT 301
174 IleGlnIlePheSerPheValAlaGluPheSerValProLeuIlePh 190
302 ATACGATTTTCTCTTTGTTGCTGAGTTCTCAGTGCATTTGCTTACT 351
190 eLeuPheAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArg 207
352 CTTTTCCTGCTTTTGGCTGTGATTTTCTCTGCGGAGCACACCGGC 401
207 ImeLysArgAsnThrValAlaGlySerArgValProGlyArgGlyAlaPro 223
402 AAATGAGAAACACAGTGGCCGCGACAGAGTTCTCGCAGGGGTGCACC 451
224 IleSerAlaLeuLeuSerIleLeuSerPheLeuIleLeuTyPheSerH 240
452 ATACGCGGTGCTGTCTATCTCTCTCTCTGATCCCTGATCTTCTCCCA 501
240 sCyMetIleLysValPheLeuSerSerLeuLysPheHisIleArgArg 257
502 CTGCATGATAAAGTTTTCCTCTCTCTCTTCAAGTTTCACATCAGAAGT 551
257 heilePheLeuPhePheIleLeuValIleGlyIleTyProSerGlyHis 273
552 TCATCTTCTGTTCTTCATCTCTGATGTGTATATACCTCTCGACAC 601
274 SerLeuIleLeuIleLeuGlyAsnProLysLeuLysGlnAsnAlaLys 290
602 TCTCTCATCTTAATTTTAGCAAAATCTTAATGTAAACAAATGCAAAA 651
290 sPheLeuLeuHisSerLysCysCysGln 299
652 GTTCCCTCTCCACAGTAACTGTCTGAG 679

seq_name: gb_gss4: A0316999
seq_documentation_block:
LOCUS A0316999 650 bp DNA GSS 22-DEC-1998
DEFINITION CITBI-EI-2530B7.TF CITBI-EI Homo sapiens genomic clone 2530B7, DNA
sequence.
ACCESSION A0316999
VERSION A0316999.1 GI:4040265
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venet, J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
Unpublished (1998)
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetlgr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:

```

```

http://www.tigr.org/cdb/humgen/Bac_end_search/Bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
source Location/Qualifiers
1..650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1fb="CITBI-EI"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltch Human BAC Library D"
BASE COUNT 149 a 152 c 125 g 224 t
ORIGIN

alignment_scores:
Quality: 1106.00 Length: 216
Ratio: 5.120 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-510-332-1 x A0316999
Align seg 1/1 to: A0316999 from: 1 to: 650

74 GluPheIleMetCysSerAlaAsnCysAlaIleLeuLeuPheIleAsnG 90
2 GAATTCATCATGTTGTTCTGCGAATTTGCAATTCCTTAATTAATGA 51
90 uLeuGluLeuTrpLeuAlaThrTrpLeuGlyValPheTyrcysAlaLys 107
52 ATTGGAATTTGGCTTGCACATGCTCGCGCTTCTTAATTTGTCAGAG 101
107 aAlaSerValArgHisProLeuPheIleTrpLeuLysMetArgIleSer 123
102 TTGCCAGCGTCCGTCACCCAGCTTCATCTGTTGAGATGAGATATCC 151
124 LysLeuValProTrpMetIleLeuGlySerLeuLeuTyValSerMetI 140
152 AGCGTGCCTCCATGATGATCTCGGGCTCTCGATATGATCATGAT 201
140 eCyValPheHisSerLysTyValGlyPheMetValProTyPheLeu 157
202 TTGTGTTTCCATAGCAAAATATGACAGGGTTATGCTCCATCTTCCT 251
157 rGlyPhePheSerGlnAsnAlaThrIleGlnLysGluAspThrLeu 173
252 GGAATTTTTCCTCCCAAAATGCCAATTCAAAAGAAATACACTGGCT 301
174 IleGlnIlePheSerPheValAlaGluPheSerValProLeuIlePh 190
302 ATACGATTTTCTCTTTGTTGCTGAGTTCTCAGTGCATTTGCTTACT 351
190 eLeuPheAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArg 207
352 CTTTTCCTGCTTTTGGCTGTGATTTTCTCTGCGGAGCACACCGGC 401
207 ImeLysArgAsnThrValAlaGlySerArgValProGlyArgGlyAlaPro 223
402 AAATGAGAAACACAGTGGCCGCGACAGAGTTCTCGCAGGGGTGCACC 451
224 IleSerAlaLeuLeuSerIleLeuSerPheLeuIleLeuTyPheSerH 240
452 ATACGCGGTGCTGTCTATCTCTCTCTGATCCCTGATCTTCTCCCA 501
240 sCyMetIleLysValPheLeuSerSerLeuLysPheHisIleArgArg 257
502 CTGCATGATAAAGTTTTCCTCTCTCTTCAAGTTTCACATCAGAAGT 551
257 heilePheLeuPhePheIleLeuValIleGlyIleTyProSerGlyHis 273

```

```

552 TCACCTTCTGCTCTTCATCCCTTGATGATATACCCCTTGACAC 601
274 SerleuileuileuLeuGlyAsnProlySleuLySglnAsnAlaLys 289
|||||
602 TCTCTCATCTTAATTGAAATGCAATTAATGAAACAAATGCAAAA 649
seq_name: gb-gss4:AQ277039
seq_documentation_block:
LOCUS AQ277039 310 bp DNA GSS 22-NOV-1998
DEFINITION CITR1-El-2516A7.TF CITR1-El Homo sapiens genomic clone 2516A7, DNA
sequence:
ACCESSION AQ277039
VERSION AQ277039.1 GI:3903235
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 310)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Sub,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
Unpublished (1998)
Other_GSSs: CITR1-El-2516A7.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
Location/Qualifiers
source 1..310
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CITR1-El"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelobAC11; Site_1: EcoRI; site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT 75 a 77 c 69 g 88 t 1 others
ORIGIN
alignment_scores:
Quality: 455.00 Length: 102
Ratio: 4.643 Gaps: 0
Percent Similarity: 96.078 Percent Identity: 90.196
alignment_block:
US-09-510-332-1 x AQ277039 ..
Align seg 1/1 to: AQ277039 from: 1 to: 310
124 LysLeuValProTrpMetIleLeuGlySerLeuLeuTyValSerMetI 140
|||||
3 AAGCTGCTCCCATGATGATCGGGGCTCTCTCTATATATATATATATGAT 52
|||||
140 eCyValPheHisSerLysTyValGlyPheMetValProTyPheLeuA 157
|||||
53 ACGGTCTCTCATAGCAATATGACGGGTTATGTCCTCATCTTCTCTAA 102
|||||
157 TgLySphePheSerGlnAsnAlaThrIleGlnLySglnAspThrLeuAla 173
|||||

```

```

103 GGAATTTNTTCTCCCAAAATGCCCAATTCAAAAGAAAGATACATGGCT 152
174 IlleGlnIlePheSerPheValAlaGlnPheSerValProIleuIlePhe 190
|||||
153 ATACGATTTACTCTCTTGTGTCGAGATTCATGATGCCATTTGCTTACTT 202
|||||
190 eLeuPheAlaValLeuLeuIleuIlePheSerLeuGlyArgHisThrArg 207
|||||
203 CCTTTTGGCTGATATGCTCTAGATTTATCTCTTGCGGAGACACACCGGCG 252
|||||
207 InMetArgAsnThrValAlaGlySerArgValProGlyArgGlyAlaPro 223
|||||
253 AAATGAGAAACACAGCGCGCGACAGAGGTTCTGCGAGGGGTGCACCC 302
224 IlSer 225
|||||
303 ATCAGC 308
seq_name: gb-est7:AA416581
seq_documentation_block:
LOCUS AA416581 463 bp mRNA EST 16-OCT-1997
DEFINITION zu05604.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:730974
3', mRNA sequence.
ACCESSION AA416581
VERSION AA416581.1 GI:2077515
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 463)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisler,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
Schellenderg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
,T., Waterston,R. and Wilson,R.
Washu-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41m13 fwd. Ex from Amersham
High quality sequence stop: 413.
Location/Qualifiers
source 1..463
/organism="Homo sapiens"
/db_xref="GDB:5927764"
/db_xref="taxon:9606"
/clone_lib="IMAG730974"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCACTGAGTGAAGTGGAGCGCGCCCAATTTTATTTTATTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 140 a 109 c 93 g 121 t
ORIGIN
alignment_scores:
Quality: 406.00 Length: 83

```







```

462 ATTTCATTCTCTCTTCCACCATGAGTTTGACGGGCAATCTCGC 413
23 rnsnlyllellelvalvalasnglylleaspleullelyshsargL 40
412 CAACGCCCTTCATTGTCTGTAATGTTGGATCTGTAATAAAGCAGC 363
40 ysmetalpProleuaspLeuleuSerCysleuAlaValSerargile 56
362 CCTTGACACACATGACATCGACCTGCTGTCTACATCAGCTCGCT 313
57 pheleuGlnleuPheleuPheleuValasVallellepheleu 73
312 TTCCTGCAG.....GGCCTTCGCTTTGGATGAT 281
73 eoljupheleuMetCys..... 78
280 TCAG...CTCGCTGCTTCCACGACATGAAAGCCACTGACCAACT 234
79 ..SerAlaAsnCysAlaIleleuPheleuAsnGluLeuGluLeuTrp 94
233 ACCAAGCCATCTCTCCTCTGATGATGCAACCAAGTACGCTCTCG 184
95 leuAlaMetTrpPheleuGlyValPheleuTyrCysAlaValAlaSerValAr 111
183 CTGCTGCTGCTCTCAGTCTCTCTACTGCTCAAGATGTCCTCTCTC 134
111 gHis.....ProleuPhele.....TripleuLysMetArgIleSerL 124
133 TCACACCTTTCACATCCATGACGAGCTGGCTCCAGGAGATTTCTTC 84
124 yslLeuValProTrpMetIleleuGlySerLeuLeuTyrValSerMetIle 140
83 AGATGCTTCTAGTGTCTTCTCTCTCATCTGACATGACCTGCTTGT 34
141 .....CysValPheHisSer 145
33 TGGGACCTTTTTCACATCTCCTCTCC 7
seq_name: gb_gss21:AZ416308
seq_documentation_block:
LOCUS AZ416308 656 bp DNA GSS 03-OCT-2000
DEFINITION 1M0191D04R Mouse 10kb plasmid UUCG1M library Mus musculus genomic
ACCESSION AZ416308
VERSION AZ416308.1 GI:10540321
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 656)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
Plasmid Inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std. Error: 0.00
Plate: 0191 row D column 04
Seq primer: CACACAGGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 656.

```

```

FEATURES
    source
        Location/Qualifiers
            1..656
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUCG1M0191D04"
                /clone_1lb="Mouse 10kb plasmid UUCG1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: pMD42ny; Purified genomic DNA from M.
                musculus C57BL/6J (male). Was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adaptor DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pMD42 (g1473214|gb|AF129072.1), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adaptor mouse DNA was annealed to
                adaptor vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."
BASE COUNT      226 a      117 c      119 g      194 t
ORIGIN
alignment_scores:
    Quality: 169.50      Length: 113
    Ratio: 2.230      Gaps: 1
    Percent Similarity: 67.257      Percent Identity: 38.053
alignment_block:
US-09-510-332-1 x AZ416308/rev ..
Align seg 1/1 to reverse of: AZ416308 from: 1 to: 656
188 leuIlePheleuPheAlaValleuLeuIlePheSerLeuIleTyr 204
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||
587 ATGATATGTTAATGTGATGTTCTCTGTTATCATTTCACTTGGAGACA 538
204 sTrpArgGlnMetArgAsnThrValAlaGlySerArgValProGlyArg 221
I|||||:|||||:|||||:|||||:|||||:|||||:|||||
537 TCCGAGGACAGATGAAATCAATTAATGATTCAGAGATCTTCACAG 488
221 lValAlaProIleSerAlaLeuSerIleLeuSerPheleuIleLeuTyr 237
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
487 AACTTCATGACACATTAATAAGTTTATGCTTTTATATATCCTTTT 438
238 PheSerHisCysMet...IleLysValPheLeuSerSerLeuPheHis 253
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
437 ATATTCATTTCAATGCGTATACCAATTAATGTAATTTGCTGTATATCCC 388
253 sIleArgArgPheIlePheleuPheIleValIleGlyIleTyrP 270
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
387 AGAACAACACTTGTATTATCATGTTGTTGACACTGATTCATCTATTC 338
270 roSerGlyHisSerLeuIleLeuIleGlyAsnProLysLeuLysGln 286
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
337 CCGGCTGCACATCATTATCTTAATTCAGCAACAGTGGCTGAAGCAG 288
287 AsnAlaLysLysPheleuLeuHisSerLysCysGln 299
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
287 TGCTGTATGATGATCTGCACTAATTAAGTCTGTGAG 249
seq_name: gb_gss10:AQ711250
seq_documentation_block:

```

LOCUS AQ711250 538 bp DNA GSS 13-JUL-1999  
 DEFINITION HS\_5349\_B2\_H04\_T7A RPCI-11 Human Male BAC Library Homo sapiens  
 genomic clone Plate=925 Col=8 Row=P, DNA sequence.  
 ACCESSION AQ711250  
 VERSION AQ711250.1 GI:5460566  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 538)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,D., Young,D., Zhao,S., Adams,M.D. and  
 Hood,L.  
 Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 JOURNAL 99380589  
 MEDLINE Contact: Mahairas GG, Wallace JC, Hood L  
 COMMENT High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
 or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
 http://www.htsc.washington.edu  
 Plate: 925 Row: P Column: 8  
 Seq primer: T7  
 Class: BAC ends  
 Location/Qualifiers  
 1..538  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate=925 Col=8 Row=P"  
 /clone\_1id="RPCI-11 Human Male BAC library"  
 /sex="male"  
 /note="Vector: pBACe3.6; Site.1: EcoRI; Site.2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBACe3.6 vector at EcoRI sites"  
 BASE COUNT 165 a 97 c 111 g 156 t 9 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 157.50 Length: 171  
 Ratio: 1.445 Gaps: 7  
 Percent Similarity: 63.743 Percent Identity: 32.164  
 alignment\_block:  
 US-09-510-332-1 x AQ711250/rev ..  
 Align seg 1/1 to reverse of: AQ711250 from: 1 to: 538  
 115 PhelIleuPhleuSerMetArgIleSerIleValProIrrpMetIlele 131  
 537 TTTCTTGGTGAATATCAGAGATCCCAATTAATACCTTGCTTCTCT 489  
 131 uGlySerLeuLeu...TyValSerMetIleCysValPheHisSerIys. 146  
 488 GCGAAGGCTCTGCGCTGTGACATGNCATCTGTGTGAGAGTAGAT 439  
 147 TyValactyPheMetValProTyPheLeuArgIlySpherPheSerGlnas 163  
 438 TACCTTAAATGTGGAAGAGATGACCTCAGA.....AA 404

163 pAlaThrIleGlnIySerGlnuSpThrIleuAlaIleGlnIlePheSerPheY 180  
 403 CACCACACTAAAGAGTANATCA...AAGATTAAGANATTAAGTAGAG 357  
 180 aAlaGluPheSerVal.....ProLeuIleIlePheLeuPhe 192  
 356 TGCCTTCTGTCACCTGCAATATATATTCCTAGCCATATATGTGATG 307  
 193 AlAValLeuLeuLeuIlePheSerLeuGlyArgHisThrArgGlnMetAr 209  
 306 TGCACCTTCTATGTTCATCATCTCTTACACGACCATCTAGATGCA 257  
 209 gasThrValAlaGlySerArgValProGlyArgGlyAlaProIleSerA 226  
 256 ACATGATCTCATGCTTTAGAAATGCCAACAGACAGACCATTAATG 207  
 226 lAlaLeuSerIleLeuSerPheLeuIleLeuTyPheSerHisCysMet 242  
 206 CATTAAGAACAGATTAACATTTCTTGCCTTTATTTCTTAT..... 163  
 243 lIleYValPheLeuSerSerLeuIlySpherHisIle.....ArgArgPh 257  
 162 TTGCTGCTTCATGACAAATATGACATTTAGTTTACCTTACAGAAATCA 113  
 257 eIllePheuPhePheIleuValIleGlyIleTyProSerGlyHis 274  
 112 CCAGTCTCTTATGCTGAGAGACATTAAGACAGCATATCCCTGCGCAGT 63  
 274 eTLeuIleLeu 277  
 62 CGGTATATATA 52  
 seq\_name: gb-gss15:AZ078264  
 seq\_documentation\_block:  
 LOCUS AZ078264 644 bp DNA GSS 31-MAR-2000  
 DEFINITION RPCI-23-399P19.TV RPCI-23 Mus musculus genomic clone RPCI-23-399P19  
 , DNA sequence.  
 ACCESSION AZ078264  
 VERSION AZ078264.1 GI:7371163  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 644)  
 Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret  
 ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.  
 and Fraser,C.M.  
 Mouse BAC End Sequences from Library RPCI-23  
 Unpublished (1999)  
 COMMENT Other GSSs: RPCI-23-399P19.TJ  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhaoc@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
 or from Resea ch Genetics (info@resgen.com). BAC end page:  
 http://www.tigr.org/tdb/bac\_ends/mouse/BAC\_end\_intro.html  
 Plate: 399 Row: P Column: 19  
 Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers  
 1..644  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"

```

/clone="RPCI-23-399P19"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
ECORI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      192 a      114 c      101 g      237 t
ORIGIN

alignment_scores:
  Quality: 150.00      Length: 181
  Ratio: 1.456      Gaps: 7
  Percent Similarity: 56.906      Percent Identity: 30.387

alignment_block:
US-09-510-332-1 x AZ078264 ..

Align seg 1/1 to: AZ078264 from: 1 to: 644

129 MetIleuGlySerLeuLeuTyValSerMetIleCysValPheHisSe 145
   ::::::::::::::::::::: :::::::::::::::
13 ATTGTCCTGAGGACACATCATGATTTTATTTTCATTTTATGCA 62
   ::::::::::::::::::::: :::::::::::::::
145 rlystyValaglyPheMetValProTyrPheLeuArgysPhe.....P 160
   ::::::::::::::::::::: :::::::::::::::
63 AATGATACCTAAT.....AATTATATACAAAGACAAAT 100
   ::::::::::::::::::::: :::::::::::::::
160 heSerGlnAsnAlaThrIleGlnLysGluAspThrLeuAlaIleGlnIle 176
   ::::::::::::::::::::: :::::::::::::::
101 TGGACAAACACACATTCCTGTTTATGATCTTAAGT..... 141
   ::::::::::::::::::::: :::::::::::::::
177 PheSerPheValAlaGluPheSerVal.....ProLeuLeuIlePheLe 191
   ::::::::::::::::::::: :::::::::::::::
142 ..GGTTTCTAGTCTACCATAGCCTTACAAATGGGATTCATTTCTT 188
   ::::::::::::::::::::: :::::::::::::::
191 uPheAlaVal.....LeuLeuLeuIlePheSerLeuGlu 203
   ::::::::::::::::::::: :::::::::::::::
189 TTTTATATAGTGTCTGACCTCATTTCTTTTAAATCTTCTCTTATGGA 238
   ::::::::::::::::::::: :::::::::::::::
203 rGlnStHArgGlnMetArgAsnThrValAlaGlySerArgValProGly 219
   ::::::::::::::::::::: :::::::::::::::
239 GGCACCTTAGAGAGATGAACTACAGGCGCATACATACCAAGACATAAGC 288
   ::::::::::::::::::::: :::::::::::::::
220 ArgGluAlaProIleSerAlaLeuLeuSerIleLeuSerPheLeu.... 234
   ::::::::::::::::::::: :::::::::::::::
289 ACAGAAAGCACATMAAGCTATGAAACATGATGTCATTCCTTTGTT 338
   ::::::::::::::::::::: :::::::::::::::
235 ....IleLeuTyrPheSerHisCysMetIleLeuValPheLeuSerSerL 250
   ::::::::::::::::::::: :::::::::::::::
339 CTTCATCATATATATATATATACCAACATTAATGCTTATGCGCAAGCTCA 388
   ::::::::::::::::::::: :::::::::::::::
250 euLysPheHisIleArgArgPheIlePheLeuPheIleLeuValIle 266
   ::::::::::::::::::::: :::::::::::::::
389 TTTCTGACAAAGTGGTGCACAAATTTTCTTATACSTA.....ATA 432
   ::::::::::::::::::::: :::::::::::::::
267 GlyIleTyrProSerGlyHisSerLeuIleLeuLeuGluLysAsnProLy 283
   ::::::::::::::::::::: :::::::::::::::
433 TTTCTGATTTATCTGTTCATCTTTCTTCTGTTTATGGAACACAA 482
   ::::::::::::::::::::: :::::::::::::::
283 sLeuLysGlnAsnAlaLysValPheLeuHisSerLysCys 297
   ::::::::::::::::::::: :::::::::::::::
483 ATTGAATGACATTCACAGCATGATGAGAAAGCTGTGT 525
   ::::::::::::::::::::: :::::::::::::::

seq_name: gb_est66:BE033430

seq_documentation_block: 1176 bp      mRNA      EST      07-JUN-2000
LOCUS      BE033430

```

```

DEFINITION      ME01F08 ME Mesembryanthemum crystallinum cDNA 5' similar to
POLYUBIDIGLITIN, mRNA sequence.
ACCESSION      BE033430
VERSION      BE033430.1 GI:8328439
KEYWORDS      EST.
SOURCE      common ice plant.
ORGANISM      Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;
Caryophyllales; Alzooceae; Mesembryanthemum.
REFERENCE      1 (bases 1 to 1176)
AUTHORS      Bonhert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferreira
,H., Kawasaki,S., McColough,A., Michalowski,C.B., Palacio,C.,
Scara,G., Wheeler,M. and Zepeda,G.R.
JOURNAL      Functional Genomics of Plant Stress Tolerance
COMMENT      Unpublished (2000)
CONTACT      Michalowski,C.B.
UNIVERSITY      University of Arizona
BIO SCIENCES      Bio Sciences West room 513, Tucson, AZ 85721, USA
TEL      520-621-7982
FAX      520-621-1697
EMAIL      chm@u.arizona.edu
FEATURES
  source
    1..1176
    /organism="Mesembryanthemum crystallinum"
    /db_xref="taxon:3544"
    /clone_1lb="ME"
    /tissue_type="roots"
    /dev_stage="5-6 weeks"
    /note="grown in hydroponics, stress 400 mM NaCl (in 0.5
    Hoagland's) 6 h stress"
BASE COUNT      129 a      324 c      65 g      658 t
ORIGIN

alignment_scores:
  Quality: 148.00      Length: 307
  Ratio: 0.902      Gaps: 13
  Percent Similarity: 53.420      Percent Identity: 25.081

alignment_block:
US-09-510-332-1 x BE033430 ..

Align seg 1/1 to: BE033430 from: 1 to: 1176

6 LeuIleTyrPheLeuAlaValIleGlnPheLeuGlyIlePhe 22
   ::::::::::::::::::::: :::::::::::::::
324 CTCTTCATATATTTCTATATGTTTCTTATGCTTTTCTTTCTTTCTTT 373
   ::::::::::::::::::::: :::::::::::::::
22 ethAsnGlyIleIleValAlaValAsnGlyIleAspLeuIleLysHis 39
   ::::::::::::::::::::: :::::::::::::::
374 TTCATTC..... 380
   ::::::::::::::::::::: :::::::::::::::
39 rGlysmMetAlaProLeuAspLeuLeuLeuSerCysLeuAlaVal.SerAs 55
   ::::::::::::::::::::: :::::::::::::::
381 .....CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCT 416
   ::::::::::::::::::::: :::::::::::::::
55 gIlePheLeuGlnLeuPheIlePheTyValAsnValIleValIlePheP 72
   ::::::::::::::::::::: :::::::::::::::
417 TCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 466
   ::::::::::::::::::::: :::::::::::::::
72 he.....IleGluPheIleMetCysSerAlaAsnCys 82
   ::::::::::::::::::::: :::::::::::::::
467 TTTCTATCTTATCTATATCTTTCTTATCTTTCTTTCTTTCTTTCTTT 516
   ::::::::::::::::::::: :::::::::::::::
83 AlaIleLeuPheLeuHisAsnGlyLeuGlyLeuTyP.....Ile 95
   ::::::::::::::::::::: :::::::::::::::
517 TTTCTCTCTCTCTTTCTATCTTTCTTTCTTTCTTTCTTTCTTTCTTT 566
   ::::::::::::::::::::: :::::::::::::::
95 uAlaThrTrpLeuGlyValPheTyCysAlaLysValAlaSerValArgH 112
   ::::::::::::::::::::: :::::::::::::::
567 TTTCTACTTTCTTTCTTTCTTTCTTT.....ATCTCTC 598
   ::::::::::::::::::::: :::::::::::::::

```



TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LNC530 row: 1 column: 02  
High quality sequence stop: 54.  
Location/Qualifiers

## FEATURES

source  
1. 1544  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3840529"  
/clone\_lib="NIH\_MGC-21"  
/issue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: placenta; Vector: pORF7; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGACGAG(G). Size-selected for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 982 a 281 c 222 g 59 t  
ORIGIN

alignment\_scores:  
Quality: 141.50 Length: 303  
Ratio: 0.963 Gaps: 18  
Percent Similarity: 48.515 Percent Identity: 26.403

## alignment\_block:

US-09-510-332-1 x BE734359/rev ..

Align seg 1/1 to reverse of: BE734359 from: 1 to: 1544

```

6 leuilellelrrpheleuAlaValIleGlnpheleuGlyIleph 22
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1148 CTTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1099
22 eThrasnGlyIleleValIleValIleAsnGlyIleAspLeuIleYshSA 39
|:|||||:|||||:|||||:|||||:|||||:|||||:
1098 CCTT.....TTGGCTTTCTTGTG..... 1080
39 rGlyMetAlaProLeuAspLeuLeuSerCysLeuAlaVal.SerAr 55
|||||:|||||:|||||:|||||:|||||:|||||:
1079 .....CTCTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1050
55 gIlePheLeuGlnLeuPheIlePheIlePheIlePheIlePhe 72
|||||:|||||:|||||:|||||:|||||:|||||:
1049 TGTCTTTGTTGGTGTCTTCTGCTTGTCTGCTGCTGCTGCTT 1000
72 heileGluPheIleMetCysSerAlaAsnCysAlaIleLeuLeuPhe 88
|||||:|||||:|||||:|||||:|||||:|||||:
999 TTTTCTGCTTTTCTT.....TTTTCGCTCTTTTGT 968
89 AsnGlnLeuGlnLeuTrpLeuAlaThrTrpLeuGlyValPheIle 105
|||||:|||||:|||||:|||||:|||||:|||||:
967 GGTGTCTTTTCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 918
105 alyValAlaSerValArgHisProLeuPheIleTrpLeuMetArg 122
|||||:|||||:|||||:|||||:|||||:|||||:
917 TCTGCTTCT.....TTTTCGCTTGTCTTCTGCTGG 886
122 leSerLysLeuValProTrpMetIleLeuGlySerLeuLeuIle 138
|||:|||||:|||||:|||||:|||||:|||||:

```

```

885 TTTGCTGCTGCTCTTTTCTTCTTTGGGCTTTCTTTCTGCTGT 836
139 MetLeuCysValPheHisSerLysTrpAlaGlyPheMetValPro 155
|||:|||||:|||||:|||||:|||||:|||||:
835 GTGCTGTGTG.....TCTCTCTGTTGGTGGTGGTGGTGGT 798
155 eleuArgLysPhePheSerGlnAsnAlaThrIleGlnLysGlnAsp 172
|||||:|||||:|||||:|||||:|||||:|||||:
797 TAGCTGTTTCTTTTCT..... 779
172 euAlaIleGlnIlePheSerPheValAlaGluPheSerValPro 188
|||||:|||||:|||||:|||||:|||||:|||||:
778 .....TGTTTTTTTCTGCTG.....TTCCTGCTGCTTGT 743
188 uIlePheLeuPheAlaValLeuLeuIlePheSerLeuGlyArgHis 205
|||||:|||||:|||||:|||||:|||||:|||||:
742 TTTGTTTTGTTTGTCTGCTGCTGCTGCTTCTTTTCGCTT..... 703
205 hrArgGlnMetArgAsnThrValAlaGlySerArgValProGly 221
|||||:|||||:|||||:|||||:|||||:
702 .....TTGCTTCT..... 694
222 AlaProIleSerAlaLeu.....LeuSerIleLeuSerPhe 234
|||:|||||:|||||:|||||:|||||:|||||:
693 .....CTTCTTTTGTGTTGTTCTGCTGCTGCTGCTGTTT 653
234 uIleLeuTrpPheSer.....HisCysMetIleLysValPheLeu 247
|||||:|||||:|||||:|||||:|||||:|||||:
652 TGTGTTTTTGTGTTGTTCTGCTGCTGCTGTTGTTGTTGTTTCT 603
248 SerSerLeuLysPheHisIleArgArgPheIlePheLeuPheIle 264
|||:|||||:|||||:|||||:|||||:|||||:
602 TTTGCTGTGTTT.....TTTGTCTTCTGCTGCTGTTTGTCT 556
264 uValIleGlyIleTrpProSerGlyHisSerLeuIleLeuGlyA 281
|||||:|||||:|||||:|||||:|||||:|||||:
555 GATTGCTTCTCTCG.....TCTCTTTTGTGTTGTTGTTT 518
281 snProLysLeuLysGlnAsnAlaLysPheLeuLeuHisSerLys 297
|||||:|||||:|||||:|||||:|||||:|||||:
517 TTTGTTTTTGTG.....TTTGTGTTGTTGTTGCTGCTGCT 483
298 Cys 298
|||:|||||:|||||:|||||:|||||:|||||:
482 TGT 480

```

seq\_name: gb\_est79:BF037592

seq\_documentation\_block:  
LOCUS BF037592 1502 bp mRNA EST 10-OCT-2000  
DEFINITION 601461160F1 NIH\_MGC\_66 Homo sapiens cDNA clone IMAGE:3864615 5',  
mRNA sequence.  
ACCESSION BF037592  
VERSION BF037592.1 GI:10745870  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1502)  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Unpublished (1999)  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: DCMB/DPF

Plate: LLCM593 row: 9 column: 16  
High quality sequence stop: 59.

## FEATURES

Location/Qualifiers  
1. 1502

Source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3864615"  
/clone\_lib="NIH\_MGC\_66"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: ovary; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies."

BASE COUNT 666 a 301 c 405 g 130 t

ORIGIN

## alignment\_scores:

Quality: 135.50 Length: 353  
Ratio: 0.852 Gaps: 17  
Percent Similarity: 45.042 Percent Identity: 22.380

## alignment\_block:

US-09-510-332-1 x BF037592/rev ..

Align seg 1/1 to reverse of: BF037592 from: 1 to: 1502

```

11 LeuLeuAlaValIleGlnPheLeuLeuGlyIlePheThrAsnGlyIle11 27
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1353 TTGCTCTGTCGTCATGCTTGTCTGGCG..... 1324
27 eValValValAsnGlyIleAspLeuIleLysHisArgLysMetAlaProL 44
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1323 .....GCTCGTC 1317
44 euAspLeuLeuLeuSerCysLeuAlaValSerArgIlePheLeuGlnLeu 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1316 TGTTCCTGCTGTCGTCGTCCTCCGCTCTCCGCTTCTCCGCTTGTCT 1267
61 PheIlePheTyrValAsnValIleValIlePhe..... 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1266 TTCTGCTTTTCTCTGCTGCTGCTGCTGTTGCTGTTTCCGCTGCTT 1217
72 PheIleGluPhe..... 75
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1216 GTTCGTCGTTTCTTGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1167
76 .....IleMetCysSer.....AlaAsnCysAlaIleLeu 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1166 TCCGGGGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1117
86 LeuPheIleAsnGluLeuGluLeuTyrPheLeuAlaThrTyrPheGlyVal 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1116 TTGTCGTG.....GTTTCCTTGTGGCTTCTCTCTCTTTTCTTGT 1073
102 eTyrCysAlaLysValAlaSerValArgHisProLeu..... 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1072 GTGCTCTGCTCTGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1023
115 ..PheIleTyrPheLeuLysMetArgIleSerLysLeuValProTyrPhe 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1022 CTCCTCCCCCCTCCGTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTC 973
131 LeuGlySerLeuLeuTyrValSerMetIleCysValPheHisSerLysTyr 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
972 CTCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 934
147 rAlaGlyPheMetValProTyrPheLeuArgLysPhePheSerGlnAsn 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
933 .....CTTTTGTCTC.....TTCTTTTCT..... 913
164 lathrIleGlnLysGluAspThrLeu.....Ala 173

```

```

912 .....GTCCTCGCCGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 868
174 IleGlnIlePheSerPheValAlaIlePheSerValProLeuLeuIlePhe 190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
867 CTCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 818
190 eLeuPheAlaValLeuLeuLeuIlePheSerLeuGlyLysHisThrArg 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
817 TCTGTC.....CTGTTATTTTCTTCTT..... 793
207 lMetArgAsnThrValAlaGlySerArgValProGlyArgValAlaPro 223
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
792 .....CTTGGCTCGTCGTCCT.....GTCCCG 769
224 lIleSerAlaLeuLeuSerIleLeuSerPhe..... 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
768 CTTCCTGCTGTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 719
234 .....LeuIleLeuTyrPheSerHisCysMetIleLysValPheLeu 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
718 TCTGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 669
248 eSerLeuLysPheHisIleArgArgPheIle.....PheLeu 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
668 CGCTCTTGTGTTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 619
261 PhePheIleLeuValIleGlyIleTyrProSerGlyHisSer..... 274
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
618 CTTCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 569
275 .....LeuIleLeu 277
568 CTTCCTCTCTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 519
277 euIleLeuGlyAsnProLysLeu.....LysGlnAsnAlaLysLys 290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
518 TGTCTTACTGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 469
291 PheLeuLeu 293
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
468 TTCCTTCTCTG 460
seq_name: gb_est68:BE250598
seq_documentation_block:
LOCUS BE250598 1652 bp mRNA EST 13-JUL-2000
DEFINITION 600943379F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960059 5',
mRNA sequence.
ACCESSION BE250598
VERSION BE250598.1 GI:9120712
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1652)
REFERENCE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: Image.lnl.gov
Plate: LLCM52 row: m column: 20
High quality sequence stop: 148.
Location/Qualifiers
1. 1652
FEATURES
source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2960059"
/clone_lib="NH_MGC-17"
/tissue_type="rhabdomyosarcoma"
/lab_host="Dh10B (phage-resistant)"
/notice="Organ: muscle; Vector: pOTB7. Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(c). Size-selected >500bp Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 800 a 279 c 458 g 115 t

ORIGIN

Alignment\_scores:

Quality:	135.50	Length:	304
Ratio:	0.948	Gaps:	11
Percent Similarity:	47.039	Percent Identity:	24.671

Alignment\_block:

US-09-510-332-1 x BE250598/rev ..

Align seg 1/1 to reverse of: BE250598 from: 1 to: 1652

```

6 LeuLeileIeTyrrPheLeuLeuAlaValIleGlnPheLeuLeuGlyIlePh 22
   ::::::::::::::::::::
1329 GTTATATATATACATCATCGTTCGTTCGTTCATTCATTCGTTCGTTCGTTCG 1280
   ::::::::::::::::::::

22 eThrasnGlyIleIleValAlaValAsnGlyIleAspLeuIleLeuHisA 39
   ::::::::::::::::::::
1279 CAGGGGTGGCGTTCGTTCATATCTGTCGTTCCTTATTCGTTCGTTC 1230
   ::::::::::::::::::::

39 rGlysmetAlaProLeuAspLeuLeuSerCysLeuAlaValSerArg 55
   ::::::::::::::::::::
1229 CGCATTCCTTCCTTCGTTCGTTCGTTCATTCGTTCGTTCGTTCGTTC 1190
   ::::::::::::::::::::

56 IlePheLeuGlnLeuPheIlePheTyrrValAsnValIleValIlePhePh 72
   ::::::::::::::::::::
1189 GCTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTC 1140
   ::::::::::::::::::::

72 eIleGlnPheIleMetCysSerAlaAsnCysAlaIleLeuLeuPheIleA 89
   ::::::::::::::::::::
1139 TATTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTC 1094
   ::::::::::::::::::::

89 snGluLeuGlnLeuTyrrPheLeuAlaThrTrpLeuGlyValPheTyrrCysAla 105
   ::::::::::::::::::::
1093 .....TTCATTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTC 1082
   ::::::::::::::::::::

106 LysValAlaIleSerValArgHisProLeuPheIleTyrrPheLysMetArgIl 122
   ::::::::::::::::::::
1081 CTTTCATTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTC 1073
   ::::::::::::::::::::

122 eSerLysLeuValProTyrrMetIleLeuGlySerLeuLeuTyrrValSerM 139
   ::::::::::::::::::::
1072 .....CTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTC 1030
   ::::::::::::::::::::

139 eTilLeuValPheHisSerLys.....TyrrAlaGly 149
   ::::::::::::::::::::
1029 GTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTC 980
   ::::::::::::::::::::

150 PheMetValProTyrrPheLeuArgLysPhe.....PheSe 161
   ::::::::::::::::::::
979 TTTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTC 930
   ::::::::::::::::::::

161 rGlnAsnAlaThrIleGlnLysGluAspThrLeuAlaIleGlnIlePheS 178
   ::::::::::::::::::::
929 TGCTGCTGCTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTCGTTC 880
   ::::::::::::::::::::

178 eTrpPheValAlaGlnPheSerValProLeuLeuIlePheLeuPheAlaVal 194
   ::::::::::::::::::::

```



---

**THIS PAGE BLANK (USPTO)**

OM of: US-09-510-332-1 to: GenEmbl:\* out-format : pfs  
Date: Mar 15, 2001 1:47 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL-frame+ pzn.model -DEV-xlp  
-O/cg2.1/USPTO.spool/US09510332/runat.06032001.093637.13632/app-query.fasta.1.359  
-DB-GenEmbl -QEMT-fastap -SUFFIX-rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPTOL=0.000 -LOOPTEXT=0.000 -GAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELTEXT=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -QEMT=pfis -NOR=ext -MINLEN=0  
-MAXLEN=200000000 -USPR=US09510332.@CGNL1.14621 -NCPU=6  
-ICPU=3 -LONGLOG -NO\_XLPPY -WAIT -THREADS=1

## Search information block:

Query: US-09-510-332-1  
Query length: 299  
Database: GenEmbl:\*  
Database sequences: 1118133  
Database length: -1736092196  
Search time (sec): 1081.220000

Sequence	Strid	Orig	ZScore	Escore	Len	Documentation
gb.pr4:AF227129	1521.00	1807.51	1.6e-92	900	1	AF227129 Homo sapiens candidate
gb.pr1:AC003015	1521.00	1765.35	1.32953	132953	1	AC003015 Human BAC clone GSI
gb.hg13:AC026787	1516.00	1758.77	8.3e-90	143719	1	AC026787 Homo sapiens chrom
gb.hg13:AC034214	1170.50	1349.58	5.0e-67	141747	1	AC034214 Homo sapiens chrom
gb.ro:AF227149	782.00	931.63	9.8e-44	1008	1	AF227149 Mus musculus candidat
gb.ro:AF227140	767.00	913.87	9.6e-43	1008	1	AF227140 Rattus norvegicus cand
gb.hg4:AC013470	438.50	480.16	1.2e-18	180438	1	AC013470 Homo sapiens chrom
gb.pr2:AC005541	434.50	480.56	1.3e-18	110626	1	AC005541 Homo sapiens clone
gb.pr1:AF227130	405.00	448.11	6.7e-19	951	1	AF227130 Homo sapiens candidate
gb.pr1:AC004979	405.00	448.11	8.4e-17	82419	1	AC004979 Homo sapiens PAC cld
gb.pr4:AF227133	376.00	451.63	5.3e-17	930	1	AF227133 Homo sapiens candidate
gb.hg15:AC016145	376.00	451.39	5.5e-17	957	1	AF227133 Homo sapiens candidate
gb.pr2:AC006518	376.00	407.49	1.5e-14	173735	1	AC006518 Homo sapiens 12p13
gb.pr4:AF227135	371.50	440.22	1.1e-16	939	1	AF227135 Homo sapiens candidate
gb.hg16:AC073342	371.50	402.43	2.9e-14	168200	1	AC073342 Homo sapiens chrom
gb.hg15:AC068649	371.50	402.23	3.0e-14	172167	1	AC068649 Homo sapiens chrom
gb.hg10:AC024156	371.50	401.67	3.2e-14	184034	1	AC024156 Homo sapiens chrom
gb.ro:AF240766	369.00	443.68	1.5e-16	894	1	AF240766 Rattus norvegicus tast
gb.hg7:AC020619	366.00	392.35	1.1e-13	256629	1	AC020619 Mus musculus clone
gb.hg10:AC024164	364.00	390.19	1.4e-13	250463	1	AC024164 Homo sapiens chrom
gb.pr4:AF227136	358.00	430.38	8.2e-16	924	1	AF227136 Homo sapiens candidate
gb.hg15:AC068649	358.00	386.25	2.3e-13	172167	1	AC068649 Homo sapiens chrom
gb.pr4:AF227137	356.00	428.12	1.1e-15	912	1	AF227137 Homo sapiens candidate
gb.ro:AF240768	355.50	427.32	1.2e-15	927	1	AF240768 Rattus norvegicus tast
gb.hg16:AC073342	350.00	376.98	7.7e-13	168200	1	AC073342 Homo sapiens chrom
gb.hg10:AC024156	350.00	376.22	8.5e-13	184034	1	AC024156 Homo sapiens chrom
gb.hg16:AC073264	346.50	371.37	1.6e-12	200128	1	AC073264 Homo sapiens chrom
gb.hg17:AC018851	346.50	371.95	1.6e-12	162380	1	AC018851 Homo sapiens clone
gb.hg18:AC079625	345.50	371.04	1.6e-12	180736	1	AC079625 Homo sapiens chrom
gb.pr3:AC010176	343.50	367.73	2.5e-12	202103	1	AC010176 Homo sapiens 12 BAC
gb.ro:AF227145	339.50	408.28	1.4e-14	945	1	AF227145 Rattus norvegicus cand
gb.hg7:AC018851	337.00	361.88	5.3e-12	162380	1	AC018851 Homo sapiens clone
gb.hg7:AC018630	337.00	354.85	1.3e-11	213227	1	AC018630 Homo sapiens chrom
gb.ro:AF227144	331.50	399.06	4.5e-14	918	1	AF227144 Rattus norvegicus cand
gb.hg3:AC011654	327.50	350.34	2.3e-11	168197	1	AC011654 Homo sapiens clone
gb.hg17:AC073647	325.50	347.73	3.3e-11	173000	1	AC073647 Homo sapiens chrom
gb.ro:AF227146	317.50	382.40	3.8e-13	927	1	AF227142 Rattus norvegicus cand
gb.ro:AF227142	317.00	381.78	4.2e-13	930	1	AF227146 Rattus norvegicus cand
gb.ro:AF227132	315.50	380.28	5.0e-13	900	1	AF227132 Homo sapiens candidate
gb.ro:AF240765	315.50	379.38	5.7e-13	1002	1	AF240765 Rattus norvegicus tast

seq\_name: gb.pr4:AF227129  
seq\_documentation\_block:  
LOCUS AF227129 900 bp DNA PRI 18-MAR-2000  
DEFINITION Homo sapiens candidate taste receptor T2R1 gene, complete cds.  
ACCESSION AF227129  
VERSION AF227129.1 GI:7262604  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 900)  
REFERENCE  
AUTHORS Adler,E., Hoon,M.A., Mueller,K.L., Chandrasekar,J., Ryba,N.J. and  
Zuker,C.S.  
A novel family of mammalian taste receptors  
JOURNAL Cell 100 (5), 693-702 (2000)  
MEDLINE 20222571  
REFERENCE 2 (bases 1 to 900)  
AUTHORS Adler,E., Hoon,M.A., Mueller,K.L., Chandrasekar,J., Ryba,N.J. and  
Zuker,C.S.  
Direct Submission  
Submitted (21-JAN-2000) NIDCR, NIH, Bldg. 10 Room 1N106, 10 Center  
Drive, Bethesda, MD 20892, USA  
Location/Qualifiers  
FEATURES  
source  
1..900  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5p15.2"  
1..19  
1..900  
/note="G protein-coupled receptor; PCR-derived sequence;  
similar to an interval in BAC AC003015"  
/codon\_start=1  
/product="candidate taste receptor T2R1"  
/protein\_id="AF43902.1"  
/translation="MLESILIVFLAVIOFLGIFNGIIVVNCIDILIKRRKAPL  
/db\_xref="GI:7262605"  
/translation="MLESILIVFLAVIOFLGIFNGIIVVNCIDILIKRRKAPL  
DLSLCASRIFLOLEIFVNVIVFIEFIEISANCAILFIEINLEMLATWLVF  
YCKVAVSRHPIELTLMKRTSKVPMNIGSLIYVNCIVFSGYKGFVPLRFKEF  
SQNRATIKETPLAIOFSPFAESVPLIFLIFLIVSLIGRHTROMNTVAGSRVP  
GRAPISALSTLSFLIFVSHCMIVFLSSLFHRRIFLFFILVIGIYPSGSHLI  
LIGNPRLKONARKFLHSKCO"  
1..>900  
/product="candidate taste receptor T2R1"  
primer\_bind 881..900  
BASE COUNT 203 a 172 g 322 t  
ORIGIN

alignment\_scores:  
Quality: 1521.00 Length: 299  
Ratio: 5.087 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-510-332-1 x AF227129  
Align seg 1/1 to: AF227129 from: 1 to: 900

1 MetLeuGluSerHisLeuIleIleTyrrPheLeuLeuAlaValIleGlnh 17  
|||||  
1 ATCTGAGAGCTCACCTCATTCATCTTTCTTCGACGTGATCAAT 50  
1 eteLeuGlyIleIlePheThrAsnGlyIleIleValIleValAsnGlyIleA 34  
|||||  
51 TCTTCTTGAGATTTCACCAATGACATCATTTGTGTGTGTAATGACATTG 100

```

34  apLeuIleuYshIaRglYsMeLaIaProLeuAspLeuLeuSerCys 50
|||||
101 ACTTATCATGACAGCAAAAAAGTGGCTCGCGTGATCTCTCTTCTGT 150
51  LeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyValAsnVa 67
|||||
151 CTGGCAGATTCTAGAAATTTTCTCGCAGTTGTTTCATCTCTACGTTAAATGT 200
67  ILeValIlePhePheIleGluPheIleMeCysSerAlaAsnCysAlaI 84
|||||
201 GATTGTATCTCTCTTCATAGAAATTCATCATGTGTCTCGAAATGTGCAA 250
84  LeLeuLeuPheIleAsnGlnLeuGlnLeuTrpLeuAlaIleThrTrpLeuGly 100
|||||
251 TTCTCTTTATTTATTAATTAATGAACTTGGCTTGGCACATGGCTCGGC 300
101 ValPheTyCysAlaValAlaSerValArgHisProLeuPheIleTr 117
|||||
301 GTTTTCTATTGTGGCCAGGTTGCCAGCGTCCGTCACCCACTCTTCATCTG 350
117  PLeuYsMeLaRglIleSerLysLeuValProTrpMetIleLeuGlySerL 134
|||||
351 GTTGAAGATGAGGATATCCAAAGCTGGTCCATGATGATCTCGGGCTCTC 400
134  euLeuTyValSerMetIleCysValPheHisSerLysTyValAglyPhe 150
|||||
401 TGCTATATGATCTATGATTTGTGTTTCCATACAAATATGACGGGTTT 450
151  MetValProTyRPhLeuArgLysPhePheSerGlnAsnAlaThrIleG 167
|||||
451 ATGGTCCCATCTTCTTAGGAAATTTTCTCCAAAAATGCCACAAATTCA 500
167  nLysGlnAspThrLeuAlaIleGlnIlePheSerPheValAlaGlyPheS 184
|||||
501 AAAAGAAGATACACTGGGTATACAGATTTTCTTTGTTGTTCTGAGTTCT 550
184  erValProLeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePheSer 200
|||||
551 CAGTGCCCATTTCTTATCTTCTCTTTTGTGTTTGTCTGATTTTCTCT 600
201  LeuGlyArgHisThrArgGlnMetArgAsnThrValAlaGlySerArgVa 217
|||||
601 CTGGGAGGACACCCCGCAATGACAAACACAGTGGCGGACGACGAGGT 650
217  lProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerPheL 234
|||||
651 TCTTGCGAGGCGGTGCACCCATCAGCGGTTGCTGTATCTGTCTTCC 700
234  euIleLeuTyRPhSerHisCysMetIleLysValIlePheLeuSerLeu 250
|||||
701 TGATCTCTTACTTCTCCACTGCAATGATAAAAGTTTCTCTCTCTCTA 750
251  LysPheHisIleArgArgPheIlePheLeuPhePheIleLeuValIleG 267
|||||
751 AAGTTTACATCAGACAGGTTTCATCTTCTGTCTTTCATCTCTGATTCG 800
267  yIleTyRProSerGlyHisSerLeuIleLeuGlyAsnProLysL 284
|||||
801 TATATACCTTCTTGACACCTCTCATCTTAATTTTAGGAAATCTTAAT 850
284  euLysGlnAsnAlaValLysPheLeuLeuHisSerLysCysGlyL 299
|||||
851 TGAACCAAAATGCAAAAAGTTCTCTCTCCACAGTAAGTGTCTGTACG 897

```

seq\_name: gb-prl:AC003015

seq\_documentation\_block:  
LOCUS AC003015 132953 bp DNA PRI 04-FEB-2000  
DEFINITION Human BAC clone GSI-113H23 from Sp15.2, complete sequence.  
ACCESSION AC003015  
VERSION AC003015.1 GI:2547255  
KEYWORDS HTG.

SOURCE  
ORGANISM human.  
REFERENCE  
AUTHORS Homo sapiens  
TITLE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
JOURNAL Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE 1 (bases 1 to 132953)  
AUTHORS Mamsley, P., Kramer, J., Elliott, G. and O'Brien, D.  
JOURNAL The sequence of H. sapiens BAC clone GSI-113H23.  
TITLE 2 (bases 1 to 132953)  
AUTHORS Unpublished  
JOURNAL Waterston, R.  
TITLE Direct Submission  
REFERENCE Submitted (20-OCT-1997) Department of Genetics, Washington  
AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
TITLE 3 (bases 1 to 132953)  
AUTHORS Waterston, R.  
JOURNAL Direct Submission  
REFERENCE Submitted (03-FEB-2000) Department of Genetics, Washington  
AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
TITLE 4 (bases 1 to 132953)  
AUTHORS Waterston, R.  
JOURNAL Direct Submission  
REFERENCE Submitted (04-FEB-2000) Department of Genetics, Washington  
AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
TITLE Genome Center  
COMMENT Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: saplens@wustl.wustl.edu  
Summary Statistics  
Center project name: H\_GSI113H23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University School of Medicine, St. Louis MO.

Mapping information for this clone was also provided by Dr. Michael Lovett, Departments of Otorhinolaryngology, Molecular Biology and Oncology, University of Texas Southwestern Medical Center, Dallas TX.

SOURCE INFORMATION:  
This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).  
Cell line: lymphoblastoid

Haplotypes: two  
VECTORS: pBel0BAC  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is GSI-330J10, 200 bp overlap. Actual start of this clone is at base position 1 of GSI-113H23; actual end is at 132953 of GSI-113H23.

This clone contains STS HSC022YA5 (NID:91235481) and HS268ZD9 (NID:9454585).

FEATURES  
Source  
This clone contains polymorphisms with GSI-330J10.  
Location/Qualifiers  
1..132953

	/organism="Homo sapiens"	/rpt_family="L1"	.36880)
	/db_xref="taxon:9606"	/complement(36832.	
	/chromosome="5"	/rpt_family="L1"	
	/clone="GSI-113h23"	repeat_region	
	/clone_lib="GSBAC1"	repeat_region	
	/map="5p15.2"	repeat_region	
	complement(20..194)	repeat_region	
	/rpt_family="L1"	repeat_region	
	195..358	repeat_region	
	/rpt_family="ALU"	repeat_region	
	complement(359..853)	repeat_region	
	/rpt_family="L1"	repeat_region	
	2903..2933	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(3640..3667)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(3764..3788)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(3820..4112)	repeat_region	
	/rpt_family="ALU"	repeat_region	
	complement(4322..4351)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(4444..4464)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(5046..5122)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(5527..6871)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(6872..6983)	repeat_region	
	/rpt_family="ALU"	repeat_region	
	complement(6985..7389)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(7459..8612)	repeat_region	
	/rpt_family="L1"	repeat_region	
	8604..9143	repeat_region	
	/rpt_family="L1"	repeat_region	
	9170..9458	repeat_region	
	/rpt_family="ALU"	repeat_region	
	9481..9546	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(9591..9823)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(9854..9961)	repeat_region	
	/rpt_family="L1"	repeat_region	
	11262..11306	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(11563..14264)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(14607..17154)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(17863..17902)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(21565..21588)	repeat_region	
	/rpt_family="L1"	repeat_region	
	24509..24814	repeat_region	
	/rpt_family="ALU"	repeat_region	
	complement(26967..27276)	repeat_region	
	/rpt_family="L1"	repeat_region	
	29009..29051	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(29140..29165)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(29892..30124)	repeat_region	
	/rpt_family="L1"	repeat_region	
	complement(30225..30258)	repeat_region	
	/rpt_family="L1"	repeat_region	
	31191..31565	repeat_region	
	/rpt_family="L1"	repeat_region	
	31649..32016	repeat_region	
	/rpt_family="L1"	repeat_region	
	33991..34027	repeat_region	
	/rpt_family="L1"	repeat_region	
	35156..35179	repeat_region	

repeat_region	/rpt_family="L1"	complement(36832.	.36880)
repeat_region	/rpt_family="L1"	39878..39924	
repeat_region	/rpt_family="L1"	complement(41297.	.41751)
repeat_region	/rpt_family="MER"	complement(44442.	.44621)
repeat_region	/rpt_family="MER"	complement(44871.	.44963)
repeat_region	/rpt_family="MER"	complement(45153.	.45205)
repeat_region	/rpt_family="MER"	complement(45255.	.45761)
repeat_region	/rpt_family="MER"	complement(45830.	.45882)
repeat_region	/rpt_family="MER"	complement(45907.	.46229)
repeat_region	/rpt_family="MER"	46840..51590	
repeat_region	/rpt_family="L1"	51617..51660	
repeat_region	/rpt_family="L1"	52068..52252	
repeat_region	/rpt_family="MER"	52278..52459	
repeat_region	/rpt_family="MER"	complement(52251.	.52521)
repeat_region	/rpt_family="MER"	53342..53635	
repeat_region	/rpt_family="ALU"	54273..54297	
repeat_region	/rpt_family="L1"	58186..58479	
repeat_region	/rpt_family="ALU"	complement(58482.	.58529)
repeat_region	/rpt_family="L1"	58482..58529	
repeat_region	/rpt_family="L1"	complement(59606.	.59897)
repeat_region	/rpt_family="ALU"	61733..62140	
repeat_region	/rpt_family="L1"	62686..62745	
repeat_region	/rpt_family="L1"	complement(63757.	.63785)
repeat_region	/rpt_family="L1"	complement(63786.	.64077)
repeat_region	/rpt_family="ALU"		

alignment_scores:	Quality: 1521.00	Length: 299
	Ratio: 5.087	Gaps: 0
Percent Similarity: 100.000	Percent Identity: 100.000	

alignment_block:	US-09-510-332-1 x AC003015	..
Align seg 1/1	to: AC003015	from: 1 to: 132953

1 MetLeuGlusErHisLeuIleIleIyRheuleuLeuAlaValIleGlnPh	17
54851 AAGCTAGACTTCACCCTCATTTATCAATTTCTTGTGAAGATACAATT	54900
17 eueLeuGIlyIIephetHrasnGIlyIleIeValaValaVasngIylea	34
54901 TCTTCTGGGATTTTCACAAATGGCATACTTGCGTGAATGCATTTG	54950
34 splEuIIeLyshIsArGIysMeTaLaProLeuAsPLeuLeuLeuSerCy	50
54951 ACTTGATCAACACAGAAAAAGCGCCCTCGATCTCCTTCTTCTTGT	55000
51 LeuAlaValaSerArgIIepheLeuGInLeuPhelIepheTyVaIaSnVa	67

```

|||||
55001 CTGGCAGTTCTAGAAATTTTCTGCAGTTGTCATCTTCTAGCTAATGT 55050
67 |||levallephelphellegluhphelmetcysserialaencysalal 84
-55051 GATGTGATCTCTCTTCATAGAAATTCATCATGTTCTGGAATTTGCAA 55100
84 |aleuleupheilleansngluenguileuptrpleualathrtrepleugly 100
55101 TTCTCTATTATTATAATGAATTGGAACCTTGCGCTGCCACATGGCTCGGC 55150
101 ValPheTyrcysAlaLysValAlaSerValAlaGHisPProLeuPheIleTr 117
|||||
55151 GTTTCCTATTGTCGCAAGGTGCGAGCGTCCGTCACCACTTCTCATCTG 55200
117 pleuTyserMetArgIleSerLyseuValProTrpMetIleLeuGlySerL 134
55201 GTTGAAGATGAGAGATATCCAAAGCTGTCCCATGATGATCCGGGGCTCTC 55250
134 euleTytrValserMetIleCysValPheHisSerLyseuValaGlyPhe 150
55251 TGCCTATATGATCTATGATTTGCTTTTCATAGCAAAATATGACAGGTTT 55300
151 MetValProTyrcPheLeuArgLysePheSerGlnAsnAlaThrIleG1 167
55301 ATGTCCTCATCTCTCTAAGGAATTTTCTCCCAAAATGCGCACAAATCA 55350
167 nLyseGlnAspThrLeuAlaIleGlnIlePheSerPheValaGlyPheS 184
55351 AAAAGAGATACACTGCGCTATACAGATTTCTCTTTTCTGCTAGTTCT 55400
184 eValPProLeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePheSer 200
55401 CAGTCCCATCTCTTAATCTTCTTTCTGTTTCTGCTCTGATTTTCTCT 55450
201 LeuGlyArgHisLethArgGlnMetArgAsnThrValaGlySerArgVa 217
55451 CTGGGAGAGCACACCGGCAATGGAACACAGTGGCGGCGAGAGG 55500
217 lProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerPheL 234
55501 TCTGGGCGAGGGGTGCACCCATCAGCGCTGCTGCTCATCTGCTCTCC 55550
234 euleLeuTyrcPheSerHisCysMetIleLyseuValPheLeuSerSerLeu 250
55551 TGAATCTTACTTCTCCACAGTGCATGATAAAAGTTTCTCTCTCTCTA 55600
251 LysPheHisLileArgArgPheIlePheLeuPhePheIleleuValIleG1 267
55601 AAGTTTCACATCAGAAAGGTCATCTTCTGTTCTTCATCCCTTGATTTGG 55650
267 YileTyrcProSerGlyHisSerLeuIleLeuIleLeuGlyAsnProLysL 284
55651 TATATACCTCTCTGACACTCTCATCTTAATTTAGGAATCTTAAT 55700
284 eulTyserGlnAsnAlaLysLysPheLeuLeuHisSerLyseuGlyGln 299
55701 TGAACAAATGCAAAAAAGTCTCTCTCCACAGTAAGTGTGTCTGAG 55747
seq_name: gb_h1g12.AC026787
seq_documentation_block:
LOCUS AC026787 143719 bp DNA HTG 18-Jul-2000
DEFINITION Homo sapiens chromosome 5 clone CTD-2143J124, WORKING DRAFT
SEQUENCE, 10 ordered pieces.
ACCESSION AC026787 GI:9256685
VERSION AC026787.3
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 143719)

```

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

DOE Joint Genome Institute.  
Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 143719)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Jul 18, 2000 this sequence version replaced gi:7711997.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----

Project Information  
Center Project Name: 682749  
Center clone name: CITB-H1\_2143J124  
-----

Summary Statistics  
Consensus quality: 136146 bases at least Q40  
Consensus quality: 141397 bases at least Q30  
Consensus quality: 142269 bases at least Q20  
Estimated insert size: 145000; pulse field gel estimation  
Estimated insert size: 143319; sum-of-contigs estimation  
Quality coverage: 5.41 in Q20 bases; pulse field gel estimation  
Quality coverage: 5.47 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1  
1769 1768: contig of 1768 bp in length  
1869 1868: gap of unknown length  
18093 18093: contig of 16225 bp in length  
18094 18193: gap of unknown length  
18194 54185: contig of 35992 bp in length  
54186 54285: gap of unknown length  
54286 84032: contig of 29747 bp in length  
84033 84132: gap of unknown length  
84133 89414: contig of 5282 bp in length  
89415 89514: gap of unknown length  
89515 97457: contig of 7943 bp in length  
97458 97557: gap of unknown length  
97558 132096: contig of 34539 bp in length  
132097 132196: gap of unknown length  
132197 134948: contig of 2752 bp in length  
134949 135048: gap of unknown length  
135049 139246: contig of 4198 bp in length  
139247 139346: gap of unknown length  
139347 143719: contig of 4373 bp in length.

FEATURES  
source  
1..143719  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2143J124"

BASE COUNT 44478 a 28860 c 28265 g 41210 t 906 others  
ORIGIN

alignment\_scores:  
Quality: 1516.00 Length: 299  
Ratio: 5.070 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.666

alignment\_block:  
us-09-510-332-1 x AC026787 ..

Align seg 1/1 to: AC026787 from: 1 to: 143719

```

1 MetLeuLusSerHisLeuIleIleTyrPheLeuLeuAlaValIleGlnPhe 17
|||||
4517 ATGCTAAGTCTACCTCATTTATCTATCTATCTTCTTCCAGTGATCAAT 4566
17 eLeuLeuGlyIlePheThrAsnGlyIleIleValIleValIleAsnGlyIle 34
|||||
4567 TCTCTTGGAGTTTTCACAAATGACATCATGTCGTGTGTAATGCAATG 4616
34 spLeuIleLysHisArgLysMetAlaProLeuAspLeuLeuSerCys 50
|||||
4617 ACTGATCAACACAGAAAATGGCTCCGGAGATCTCTCTTCTTCTGTGT 4666
51 LeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyrValAsnVa 67
|||||
4667 CTGGCAGTTTCTAGAAATTTTCTGCAGTTTCTCATCTTCACTTAATGCT 4716
67 IleValIlePhePheIleGlnPheIleMetCysSerAlaAsnCysAlaI 84
|||||
4717 GATTTTATCTTCTCATAGAAATTCATCATGTCGTGTGCAATGTCGCA 4766
84 IleLeuLeuPheIleAsnGlnLeuGlnLeuTyrPheAlaIleThrTripleu 100
|||||
4767 TTCTCTTATTTATTAATGAATGGAACTTGGCTTGGCCACATGGCTCGGC 4816
101 ValPheTyrCysAlaLysValAlaSerValArgHisProLeuPheIleTr 117
|||||
4817 GTTTTCTATTTGGCCAGAGTTCCAGCTCCATCCACCATCTTCACTCG 4866
117 PLeuLysMetArgIleSerLysLeuValProTyrMetIleLeuGlySerL 134
|||||
4867 GTTGAAGATGAGATATCCAAAGCTGCTCCATGATGATCTCGGGGCTCT 4916
134 eLeuLeuTyrAlaSerMetIleCysValPheHisSerLysTyrAlaGlyPhe 150
|||||
4917 TCTCTATATCTATGATTTGTTGTTCCATAGCAAAATATACAGAGTTT 4966
151 MetValProTyrPheLeuArgLysPhePheSerGlnAsnAlaIleThrIle 167
|||||
4967 ATGGTCCCACTACTCTCTAAGAAATTTTCTCCAAAATGCCACCAATTC 5016
167 nLysGlnAspThrLeuAlaIleGlnIlePheSerPheValAlaGlnPheS 184
|||||
5017 AAAAGAAGATACACTGCTATACAGATTTTCTTTGTTGCTGCACTTCT 5066
184 eValProLeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePheSer 200
|||||
5067 CAGTGCATTTGCTTATCTTCTTTTGGCTGTTTGGCTTGAATTTTCTCT 5116
201 LeuGlyArgHisThrArgIleMetArgAsnThrValAlaGlySerArgVa 217
|||||
5117 CTGGGAGGACACCCGCGCAAAATGAGAAACACAGTGGCGCGCAGAGGT 5166
217 LProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerPheL 234
|||||
5167 TCTTGGAGGGGTGACCCATCAGCGGCTTCTCTATCTTCTTCTTCC 5216
234 euleLeuTyrPheSerHisCysMetIleLysValIlePheLeuSerSerLeu 250
|||||
5217 TGATTCCTTACTTCTCCACTGCATGATTAAGTTTCTCTCTCTCTA 5266
251 LysPheHisIleArgArgPheIlePheLeuPhePheIleLeuValIleG 267
|||||
5267 AAGTTTACATCAGAAAGTTTCATCTTCTCTCATCTTGGATTTGG 5316
267 YIleTyrProSerGlyHisSerLeuIleLeuIleLeuGlyAsnProLysL 284
|||||
5317 TATATACCTTCTTGGACACTCTCTCATCTTAATTTAGAAATCTTAAT 5366
284 eulYsGlnAsnAlaLysLysPheLeuLeuHisSerLysCysGlyGln 299
|||||
5367 TGAACAACAAATGCAAAAAGTTCTCTCTCCACAGTAAGTGTGTGAG 5413

```

seq\_name: gb\_hcg13:AC034214

```

seq_documentation_block:
LOCUS AC034214 141747 bp DNA HTG 18-JUL-2000
DEFINITION Homo sapiens chromosome 5 clone CTD-2001E22, WORKING DRAFT
ACCESSION AC034214
VERSION AC034214.4 GI:9256725
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
AUTHORS DOE Joint Genome Institute.
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
COMMENT Submitted (05-Apr-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7712070.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 628051
Center clone name: CITB-H1_2001E22
-----
Summary Statistics
Consensus quality: 129482 bases at least Q40
Consensus quality: 137633 bases at least Q30
Consensus quality: 139230 bases at least Q20
Estimated insert size: 128750; agarose-fp estimation
Estimated insert size: 141047; sum-of-ctnigs estimation
Quality coverage: 5.08 in Q20 bases; sum-of-ctnigs estimation
Quality coverage: 4.64 in Q20 bases; sum-of-ctnigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. Gaps between the contigs
* are represented as runs of N. The order of the sizes
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 21582: contig of 21582 bp in length
* 21583 21682: gap of unknown length
* 21683 34005: contig of 12323 bp in length
* 34006 34105: gap of unknown length
* 34106 37105: contig of 3000 bp in length
* 37105 37205: gap of unknown length
* 37205 39674: contig of 2469 bp in length
* 39674 39775: gap of unknown length
* 39775 46528: contig of 6754 bp in length
* 46528 46629: gap of unknown length
* 46629 46637: contig of 6009 bp in length
* 46637 52738: gap of unknown length
* 52738 52738: gap of 4360 bp in length
* 52738 57197: gap of unknown length
* 57197 58513: contig of 1316 bp in length
* 58513 58514: gap of unknown length
* 58514 73885: contig of 15272 bp in length
* 73885 73985: gap of unknown length
* 73985 73986: contig of 18593 bp in length
* 73986 92579: gap of unknown length
* 92579 103695: contig of 11017 bp in length
* 103695 103795: gap of unknown length
* 103795 106232: contig of 2437 bp in length

```

```

* 106233 106332: gap of unknown length
* 106333 121673: contig of 15341 bp in length
* 121674 121773: gap of unknown length
* 121774 124114: contig of 2341 bp in length
* 124115 124214: gap of unknown length
* 124215 131887: contig of 7673 bp in length
* 131888 141747: gap of unknown length
* 131988 141747: contig of 9760 bp in length.
FEATURES
    source
        1. 141747
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="CFD-2001E22"
            /clone_lib="Caltech human BAC library D"
BASE COUNT 41436 a 28372 c 28509 g 41923 t 1507 others
ORIGIN
alignment_scores:
    Quality: 1170.50      Length: 302
    Ratio: 4.417          Gaps: 4
    Percent Similarity: 87.748    Percent Identity: 84.768
alignment_block:
US-09-510-332-1 x AC034214
Align seg 1/1 to: AC034214 from: 1 to: 141747
1 MetLeuGluSerHisLeuIleIleTyrPheLeuLeuAlaValIleGlnP 17
|||||
39158 ATGCTAGAGTCTCAGCTCATTTATCTTCTTCTGAGGATGACATTT 39207
17 eLeuLeuGlyIlePhePheHisGlyIleIleValIleValIleAsnGlyIleA 34
|||||
39208 TCTTCTGGGATTTTACAAATGCGATCATGTGTGCGTGAATGGCATTTG 39257
34 sPLeuIleIleHisArgIleMetAlaProLeuAspLeuLeuSerCys 50
|||||
39258 ACTGATGTCACACAGAAATGGCTCCGCGATCTCTCTTCTTCTTCTTCT 39307
51 LeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyrValIleAsnVa 67
|||||
39308 CTGGCAGTTTCTAGAAATTTTCTCAGTTGTTCATCTTCTTACGTTAATGT 39357
67 lIleValIlePhePheIleGluPheIleMetCysSerAlaAsnCysAlaI 84
|||||
39358 GATTTGTTATCTTCTTCAATGATCATGTCGTTCGCAATTTGTCGAA 39407
84 lLeuLeuPheIleAsnGlnLeuGluLeuTyrPheLeuAlaThrTrpLeuGly 100
|||||
39408 TTCTCTTATTTATTAATGAATTTGGAACTTTGGCTTGCACATGGCTGGC 39457
101 ValPheTyrCysAlaIleValIleSerValArgHisProLeuPheIleTr 117
|||||
39458 GTTTTCTATTGTGCCAAGTTGCCAGCGTCGCCACCTCTTCATCTG 39507
117 PLeuIleMetArgIleSerIleLeuValIleProTyrMetIleLeuGlySer 134
|||||
39508 GTTGAACATAGAGATATCCAAAGCTGGTCCCATGATGATGCTT. GGGCTTC 39556
134 euleuTyrValIleSerMetIleCysValIlePheHisSerIleVal. TyrAla. GlyP 150
|||||
39557 TGCATATGATGATCATGATTTGTGTTTTCATAGCAAACTATATCAGAGAT 39606
150 heMetValIleProTyrPheLeuArgIlePhePheSerGlnAsnAlaThrIle 166
|||||
39607 TTATGGTCCCATCTCTCTTAAGGAAATTTTCTCCAAATATGCCACAAT 39656
167 GlnLeuGlnSerPheLeuAlaIleGlnIlePheSerPheValIleGluP 183
|||||
39657 CAAAGAGACATACACTGCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 39703

```

```

183 eSerValProLeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePheS 200
39704 .....NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 39728
200 eLeuGlyArgHisThrArgGlnMetArgAsnThrValAlaGlySerArg 216
:::
39729 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 39778
217 ValProGly. ArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerP 233
|||||
39779 GTTCTGGACAGAGGGTGCACCAATACAGCGCTTCTCTC. ATCTGTGCT 39827
233 heLeuIleLeuTyrPheSerHisCysMetIleLeuValIlePheLeuSer 249
|||||
39828 TCTGATCTCTACTTCTCCACATGATGATAAAGTTTCTCTCTCTCT 39877
250 LeuLysPheHisIleArgArgPheIlePheLeuPhePheIleLeuValI 266
|||||
39878 CTAAAGTTTCACATCAGAGATTCATCTTCTGTCTTCATCTCTGTGAT 39927
266 eGlyIleTyrProSerGlyHisSerLeuIleLeuIleLeuGlyAsnPro 283
|||||
39928 TGTATATACCTTCTGACACCTCTCATCTTAATTTAGGAAATCTTA 39977
283 ySLeuLysGlnAsnAlaIleLysPheLeuLeuHisSerIleCysCysGln 299
|||||
39978 AATGGAACAACAAATGCAAAAGTCTCTCTCCACAGTAACTGCTGTGAC 40027
seq_name: gb-ro:AF227149
seq_documentation_block:
LOCUS AF227149 1008 bp DNA ROD 18-MAR-2000
DEFINITION Mus musculus candidate taste receptor T2R19 gene, complete cds.
ACCESSION AF227149
VERSION AF227149.1 GI:7262644
KEYWORDS
SOURCE
    house mouse.
    ORGANISM
        Mus musculus
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
        1 (bases 1 to 1008)
        Adler,E., Hoon,M.A., Mueller,K.L., Chandrasekar,J., Ryba,N.J. and
        Zuker,C.S.
        A novel family of mammalian taste receptors
        JOURNAL
            CELL 100 (6), 693-702 (2000)
        MEDLINE
            20222571
        REFERENCE
            2 (bases 1 to 1008)
            Adler,E., Hoon,M.A., Mueller,K.L., Chandrasekar,J., Ryba,N.J. P.
            and Zuker,C.S.
            Direct Submission
            Submitted (21-JAN-2000) NIDCR, NIH, Bldg. 10 Room 1N106, 10 Center
            Drive, Bethesda, MD 20892, USA
FEATURES
    source
        1. 1008
            Location/Qualifiers
            1. 1008
                /organism="Mus musculus"
                /strain="129/SvJ"
                /db_xref="taxon:10090"
                1. 1008
                    /note="G protein-coupled receptor"
                    /codon_start=1
                    /product="candidate taste receptor T2R19"
                    /protein_id="AA043922.1"
                    /db_xref="GI:7262645"
                    /translation="MDEGHLFELVVVQFLGVNLGLIVVNAIDLIMKKMAPL
                    DLFLFCATRIILQCLILEAQLGSLVHRTLFADNVTVEYIINELSLFATWLGVE
                    YCAIATIPHLFLMKRISRLVPMILASVYVYVTFIHSRETSELKQIFISF
                    SKNTIRVPAHATILSVFGLTLPILFVAVALLLSAMNHSROMKRTVVGREPSR
                    HALVASLSTSELIELSLSDMAVVLCTGCLIFGSRFAFCLLVIGMTPLSHIVLI
                    LGNFKLNRNKTFLVHCKCHCARAWTSNRLDLPVATHSANKTSKSCSEACIMP
                    S"
BASE COUNT 224 a 252 c 209 g 323 t
ORIGIN

```







FEATURES  
source

109051 121518: contig of 12468 bp in length  
\* 121519 121618: gap of unknown length  
\* 121619 133501: contig of 11883 bp in length  
\* 133502 133601: gap of unknown length  
\* 133602 151070: contig of 17469 bp in length  
\* 151071 151170: gap of unknown length  
\* 151171 180438: contig of 29268 bp in length.  
Location/Qualifiers  
1. 180438  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/clone="RP11-472F21"  
1. 1282  
/note="assembly\_name:Contig12"  
1383. 2619  
/note="assembly\_name:Contig13"  
2720. 4393  
/note="assembly\_name:Contig14"  
4494. 10415  
/note="assembly\_name:Contig16"  
10516. 17141  
/note="assembly\_name:Contig17"  
17242. 24045  
/note="assembly\_name:Contig18"  
24146. 32644  
/note="assembly\_name:Contig19"  
32745. 41130  
/note="assembly\_name:Contig20"  
41231. 49280  
/note="assembly\_name:Contig21"  
49381. 60265  
/note="assembly\_name:Contig22"  
60366. 72486  
/note="assembly\_name:Contig23"  
72587. 84443  
/note="assembly\_name:Contig24"  
84544. 95472  
/note="assembly\_name:Contig25"  
95573. 108950  
/note="assembly\_name:Contig26"  
109051. 121518  
/note="assembly\_name:Contig27  
clone\_end:5P6  
vector\_side:left"  
121619. 133501  
/note="assembly\_name:Contig28"  
133602. 151070  
/note="assembly\_name:Contig29"  
151171. 180438  
/note="assembly\_name:Contig30"

BASE COUNT 57874 a 32227 c 32003 g 56627 t 1707 others  
ORIGIN

alignment\_scores: 438.50 length: 301  
Quality: 2.118 Gaps: 7  
Ratio: 68.771 Percent Identity: 33.555  
Percent Similarity: 68.771

alignment block:  
US-09-510-332-1 x AC013470/rev ..

Align seg 1/1 to reverse of: AC013470 from: 1 to: 180438

```

7  llelelyrphleuAAlaValIleGlnPheleuLeuGlyIlePheTh 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117119 ATTCTTCATATTCATCATGTCAGACGAATTCCTTACACAGGACACAGT 117070
23 rAsnGlyIleIleValValIleAsnGlyIleAspLeuIleLysHisArgL 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117069 AAATGCATTCCTTATCATCTTAACCTGAATGAATGATCAACATAGAA 117020

```

```

40 ysmetaLapProleuAspLeuLeuSerCysLeuAlaValSerArgIle 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117019 AGCTAATGCCAATTCATCAATCTTTAATGTGCTACGACATGCTATATT 116970
57 PheLeuGlnLeuPheIlePheYrValAsnValIleValIlePhePhe.. 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116969 GCGTCGACATGGGTAAATGTAACAAGTTTTCCTGCTGCTCTTTC 116920
73 ..... IleGluPheIleMetCysSerAlaAsnGlyAlaIleLeu 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116919 ACCTCTTACGTCATAAATAATTAATGTCGACGAATGATCTCTTGGCA 116870
86 eupheIleAsnGluLeuGluLeuTrpLeuAlaThrTrpLeuGlyValPhe 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116869 TGTTTTATAGCTATATACGCTATGTTGCCACTTCCCTTCTGTATTT 116820
103 TyrCysAlaLysValAlaSerValArgHisProLeuPheIleTrpLeu 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116819 TACTGCCCTCAAGATTTCAGGCTTCACATCCTGTTTCTTCTGTTGAA 116770
119 SmetArgIleSerLysLeuValProTrpMetIleLeuGlySerLeuLeu 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116769 ATTCAGGATCCCAAGTTAATACCTTGCTGCTCTTGGAAGCCTTGG 116720
136 ..TyrValSerMetIleCysValPheHisSerLys..TyrAlaGlyPhe 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116719 CCTCTGACGATTCATCTGCTGTGTCAGGTAGATACGTAATAATGT 116670
151 tValProTrpLeuArgLysPhePheSerGlnAsnAlaThrIleGlnL 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116669 GGAAGAGGATGCCCTCAGA.....ACACCACTAATAAAA 116635
168 ysgLAspTrpLeuAlaIleGlnIlePheSerPheVal.....Ala 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116634 AGAGTAATAACAAAGATAAAGAAATTAATGTAAGCTTCTGTCACTG 116585
182 GluPheSerValProLeuLeuIlePheLeuPheAlaValLeuLeuL 198
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116584 GCATTAAATATTCCTCTAGCCATATTTGATGTGCACCTTATGTTACT 116535
198 ePheSerLeuGlyArgHisThrArgGlnMetArgAsnThrValAlaLys 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116534 CATCTCTCTTTCAAGCACACATCGATGCAACATGATCTCATGGCT 116485
215 eArgValProGlyArgGlyAlaProIleSerAlaLeuSerIleLeu 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116484 TTAGAAATGCCAACACAGAACCCCATATAATGATTAATAAACATGATA 116435
232 SerPheLeuIleLeuTrpPheSerHisCysMetIleLysValPheLeu 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116434 ACATCTTTTGGCTTCTTATTTCTTAT.....TTTGGCTGCTTCATGAC 116391
248 rSerLeuLysPheHisIleArgArgPheIlePheLeuPheIleLeu 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116390 AAATATGACATTTAGTTTACTTTACAGAAAGTCACACGTTCTTTATGCTGA 116341
265 .....ValIleGlyIleTrpProSerGlyHisSerLeuIleLeuLeu 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116340 AGACACATATAGCAGATATCCCTGTGGCAGCTCGGTATATAATCTTG 116291
280 GlyAsnProLysLeuLysGlnAsnAlaLysLysPheLeuLeuHisSer 296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116290 AGTAATTCCTTAAGTCCACACATCATTTAGAAAGAAATTCCTGCTCAADA 116241
296 s 296
116240 G 116240
seq_name: gb_pt2:AC005541

```

seq\_documentation block:  
LOCUS AC005541 110626 bp DNA 14-JAN-1999  
DEFINITION Homo sapiens clone RG044L22, complete sequence.

```

ACCESSION   AC005541
VERSION     AC005541.1
KEYWORDS    HTG;
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     1 (bases 1 to 110626)
TITLE       Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL     The sequence of Homo sapiens clone
REFERENCE   unpublished
AUTHORS     2 (bases 1 to 110626)
TITLE       Waterston, R.H.
JOURNAL     Direct Submission
REFERENCE   Submitted (25-AUG-1998) Genome Sequencing Center, Washington
AUTHORS     University School of Medicine, 4444 Forest Park Parkway, St. Louis,
TITLE       MO 63108, USA
JOURNAL     3 (bases 1 to 110626)
REFERENCE   Waterston, R.H.
AUTHORS     Direct Submission
TITLE       Submitted (14-JAN-1999) Genome Sequencing Center, Washington
JOURNAL     University School of Medicine, 4444 Forest Park Parkway, St. Louis,
REFERENCE   MO 63108, USA
COMMENT     On Jan 14, 1999 this sequence version replaced gi:3907453.
FEATURES
     source
         1..110626
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="RC044122"
BASE COUNT   35376 a 19647 c 20125 g 35478 t
ORIGIN

```

```

Alignment scores:
  Quality: 434.50      Length: 301
  Ratio: 2.099         Gaps: 7
  Percent Similarity: 68.771   Percent Identity: 33.223

```

## Alignment block:

US-09-510-332-1 x AC005541/rev ..

Align seg 1/1 to reverse of: AC005541 from: 1 to: 110626

```

7   llelelyrphelleuAlaVallelelnpheleuGlyllepeth 23
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4392 ATTCTTCATATTCATGATGTCACAGAAATTCCTCACAGGATCACAGT 4343
23 rAsnGlylleleValValAsnGlylleAspLeuileYshsArgL 40
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
4342 AAATGAGATTCTTATCATGTTTAACTGTAAATGATGATCAAAACATAGAA 4293
40 ysmetAlaProleuAspLeuLeuSerCysLeuAlaValSerArgIle 56
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4292 AGCTAAATGCCAAATTCCTTAAATGTCACATGAGGATGCTAAGATT 4243
57 pheleuGlnleuphelelyrValAsnVallellepeth 72
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4242 GGTGTGCGATGCGTGAATGTCACAAAGTTTCTCTGTTGTTCTTCC 4193
73 .....lleGluPheileMetCysSerAlaAsnCysAlaIleleuL 86
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
4192 ACTGCTTATTCACAAATTAATTTATGTCGACGAATGATGTCCTTGA 4143
86 eupheileAngluLeuGlnleuLeuPheleuAlaThrPheleuGlyValPhe 102
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
4142 TGTTTTAACTATACGCTATGCTGTTGTCACATGCTTCTGTATTT 4093
103 TyrcAlaIaYsValAlaSerValArgIlaPheleuPheleuIleu 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4092 TACGCGCTCAAGATTTCAGGCTTCACATGCTGTTTCTTGGTGA 4043
119 smetArgIleSerlyLeuValProThrMetIleleuGlySerleuLeu. 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

4042 ATTCAGATCCCAAGTTAATACCTTGGCTCTCTGCGAAGCGTTCTG 3993
136 ..TyValSerMetIleCysValPheHisSerly TyValAglyPheMe 151
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3992 CCTCTGTAGACATTCATCTGTGTCTGTCAGATGATGCTTAACGTA 3943
151 tValProTyrrPheleuArglyPhePheSerGlnAsnAlaThrIleGlnL 168
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
3942 GGAACAGATGCCCTCAGA.....AACACCACTAAATAA 3908
168 ySGLuAspThrleuAlaIleGlnIlePheSerPheVal.....Ala 181
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3907 AGAGTAAACAAAGATTAAGAAATTAAGTGAAGCTTCTTGTCACTG 3858
182 GlupheSerValProleuLeuIlePheleuPheAlaValleuLeuL 198
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
3857 GCATTAAATATTCCTCTAGCCATATTTGTGATGTCACCTTCATGTTACT 3808
198 ePheSerleuGlyArgIshThrArgGlnMetArgAsnThrValAlaGly 215
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
3807 CATCTCTTTTACAAGCACTCATGCGATGCAACATGATTCATGCT 3758
215 eArgValProGlyArgGlyAlaProIleSerAlaLeuSerIleLeu 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3757 TTAAATATGCCAACACAGAACCCATTAATGATTAATAAACAGTATA 3708
232 SerPheleuIleleTyrrPheSerHisCysMetIleLyValPheLeuSe 248
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
3707 ACATCTCTTTGCTCTTTAATTTCTTAT.....TTGCTGCTTCATGAC 3664
248 rSerleuLyPheHisIleAlrGlyrPheIlePheleuPhePheIleu. 264
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
3663 AAATATACATTTAGTTTACCTTACACAGAGTCACAGCTTCTTTATGCTGA 3614
265 ....ValIleGlyIleTyrrProSerGlyHisSerleuIleleu 279
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
3613 AGCACAATATGCGCAGCATATCCCTGCGCCATGCTTAATATATCTTG 3564
280 GlyAsnProLySerleuGlnAsnAlaLyLyPheleuLeuHisSerly 296
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
3563 AGTAAATCTAAGTTCACACATCATTTAGAAAGAAATTCCTGCTCAAAA 3514
296 s 296
3513 G 3513
seq_name: gb_ptr4:AF227130
seq_documentation_block:
LOCUS       AF227130          951 bp      DNA          PRI          18-MAR-2000
DEFINITION  Homo sapiens candidate taste receptor T2R3 gene, complete cds.
ACCESSION   AF227130
VERSION     AF227130.1   GI:7262606
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS     Adler E., Hoon M.A., Mueller K.L., Chandrasekhar J., Ryba N.J. and
           Zuker C.S.
TITLE       A novel family of mammalian taste receptors
JOURNAL     Cell 100 (6), 693-702 (2000)
MEDLINE     20222571
REFERENCE   2 (bases 1 to 951)
AUTHORS     Adler E., Hoon M.A., Mueller K.L., Chandrasekhar J., Ryba N.J.P.
           and Zuker C.S.
TITLE       Direct Submission
JOURNAL     Submitted (21-JAN-2000) NIDCR, NIH, Bldg. 10 Room 1N106, 10 Center
           Drive, Bethesda, MD 20892, USA
FEATURES
     source
         1..951
         /organism="Homo sapiens"

```



Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
----- Summary Statistics -----  
Center project name: H\_DUI154E09

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send [mailto:sgreen@nhgri.nih.gov](mailto:mailto:sgreen@nhgri.nih.gov), or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.  
The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.  
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP5-894A10, 200 bp overlap: the clone sequenced to the right is unknown. Actual start of this clone is at base position 103168 of RP5-894A10; actual end is at 82419 of RP5-1154E9.

#### FEATURES

##### source

1.82419  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q31.3-q32"  
/clone="RP5-1154E9"  
/clone\_1lb="RPCI-5"  
1.02  
repeat\_region  
/rpt\_family="Alu"  
259.800  
/rpt\_family="MER1\_type"  
repeat\_region  
801.1104  
/rpt\_family="Alu"  
1105.1132  
/rpt\_family="MER1\_type"  
repeat\_region  
1134.1434  
/rpt\_family="Alu"  
1437.1740  
/rpt\_family="L1"  
repeat\_region  
1830.1865  
/rpt\_family="MIR"  
2441.2765  
/rpt\_family="MER2\_type"  
repeat\_region  
2918.3211  
/rpt\_family="Alu"  
3226.3272  
/rpt\_family="AT-rich"  
misc\_feature  
3239.3594

misc\_feature  
3365.3878  
/note="match to EST C15294 (NID:91570001)"  
misc\_feature  
3498.3878  
/note="match to EST AA649071 (NID:92575500) ns42h02.s1"  
STS  
3510.3756  
/note="match to EST A1031578 (NID:93249790) OW0612.x1"  
/db\_xref="GI:435384"  
misc\_feature  
3587.3874  
/note="similar to EST D57847 (NID:9664469)"  
repeat\_region  
4204.4293  
/rpt\_family="L2"  
4543.5520  
/rpt\_family="Retroviral"  
5535.8698  
/rpt\_family="Retroviral"  
8789.9520  
/rpt\_family="Retroviral"  
9522.9556  
/rpt\_family="Retroviral"  
11462.11528  
/rpt\_family="AT-rich"  
11735.12045  
/rpt\_family="Alu"  
12587.12789  
/rpt\_family="MIR"  
12884.13241  
/rpt\_family="MaLR"  
13244.13907  
/rpt\_family="MaLR"  
13902.14230  
/rpt\_family="MaLR"  
14247.14415  
/rpt\_family="L2"  
14427.15098  
/rpt\_family="L2"  
15151.15195  
/rpt\_family="AT-rich"  
15213.15353  
/rpt\_family="Alu"  
15546.16058  
/rpt\_family="MER73-group"  
16279.16492  
/rpt\_family="MaLR"  
18279.18327  
/rpt\_family="purine-rich"  
18911.19019  
/rpt\_family="MIR"  
19567.19706  
/rpt\_family="MIR"  
19707.19753  
/rpt\_family="(CA)n"  
19754.19835  
/rpt\_family="MIR"  
20529.20588  
/rpt\_family="L2"  
20778.21195  
/rpt\_family="Retroviral"  
21265.21293  
/rpt\_family="(CA)n"  
21294.21896  
/rpt\_family="L2"  
21969.22393  
/rpt\_family="L2"  
22470.22492  
/rpt\_family="MIR"  
22493.22865  
/rpt\_family="MaLR"  
22866.22954  
/rpt\_family="MIR"  
23042.23082  
/rpt\_family="AT-rich"  
23385.29503  
/rpt\_family="L1"

repeat_region	33343..33635	/rpt_family="Alu"
repeat_region	34205..34235	/rpt_family="AT_rich"
repeat_region	34523..34757	/rpt_family="MaLR"
repeat_region	35125..35170	/rpt_family="AT_rich"
repeat_region	35173..35222	/rpt_family="TAGA)n"
repeat_region	35237..35263	/rpt_family="TAGA)n"
repeat_region	35862..36050	/rpt_family="MIR"
repeat_region	36544..36581	/rpt_family="AT_rich"
repeat_region	37719..37867	/rpt_family="MIR"
repeat_region	38203..38257	/rpt_family="AT_rich"
repeat_region	38259..38372	/rpt_family="AT_rich"

  

Alignment_scores:			
Quality:	405.00	Length:	313
Ratio:	2.166	Gaps:	8
Percent Similarity:	59.744	Percent Identity:	36.422

  

Alignment block:  
US-09-510-332-1 x AC004979 ..

  

Align seg 1/1 to: AC004979 from: 1 to: 82419

  

8	lIeYrPheLeuLeuAlaValIleGlnPheLeuLeuGlyIlePheThras	24
17597	GGTTCCTGATTCCTGCTGGCAGCTAGTTCACACGGGAATTCGGTCA	17646
24	nGlyIleLeuAlaValAlaSnGlyIleAspLeuLeuIleYshIsArgLysM	41
17647	TGTGTTCATGTGGTGGTCAAGTGAAGCAGGCTGTCGAAGCAGCAAGAA	17696
41	eAlaIProLeuSpleuLeuLeuSerCysLeuAlaValAspArgIlePhe	57
17697	TGCTCTTGCTGACTTCATCATCACACCCCTGGCACTCTTGAGATCAT	17746
58	LeuGlnLeuPheIlePheTyrValAsnValIleValIlePhePheIleG	74
17747	CTGCTGTGATATCTTCACTGATAGT.....TTTTTAATAGA	17784
74	uPheIleMetCysSerAlaAspCysAlaIleLeuLeu.....	86
17785	ATTCCTGCCACACACATGATTCAGGCAATAATATGCAATATTATGATG	17834
87	.....PheIleAsnGlnLeuGlnLeuTyrLeuAlaThrTyrLeu	99
17835	TTTCCGGAGCATTTACAAACCATCTGACGATTTGGCTTCCACCTGCTT	17884
100	GlyAlaPheTyrCysAlaLysValAlaSerValArgHisProLeuPheI	116
17885	GGGTGCTCTACAGCTGAAATCGCCAGTTTCTCCACACCCACATTCCT	17934
116	eTyrPLeuYsMetArgIleSerLysLeuValProTyrMetIleLeuGlyS	133
17935	CTGGCTCACTGAGAGATTTCTAGGGTGATGATGATGATGCTGGCGG	17984
133	eTyrLeuTyr.....ValSer	138
17985	CACGCTCTTATCCCTGTGTAGTACCGCATCTGTATCATGAGTTTAAG	18034
139	MetIleCysValPheHisSerLysTyrAlaGlyPheMetValProTyrPh	155
18035	CTCATTTCTGCTTTAGGGAAATTTGAGGCCACAGGAATGTGACTGAACA	18084
155	eLeuArgLysPhePheSerLinsAlaThrIleGlnLysGluAspThrL	172

```

      seq_name= gb_pr4:AF227134
seq_documentation_block:   930 bp    DNA       PRI     18-MAR-2000
LOCUS        AF227134               complete cds.
DEFINITION   Homo sapiens candidate taste receptor T2R8 gene, complete cds.
ACCESSION    AF227134
VERSION      AF227134.1 GI:7262614
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 930)
AUTHORS      Adler,E., Hoon,M.A., Mueller,K.L., Chandrashekar,J., Ryba,N.J. and
              Zuker,C.S.
TITLE        A novel family of mammalian taste receptors
JOURNAL      Cell 100 (6), 693-702 (2000)
MEDLINE      20222571
ADDITIONAL   2 (bases 1 to 930)
REFERENCES    Adler,E., Hoon,M.A., Mueller,K.L., Chandrashekar,J., Ryba,N.J.P.
              and Zuker,C.S.
DIRECT SUBMISSION Submitted (21-JAN-2000) NIDCR, NIH, Bldg. 10 Room IN106, 10 Center
DRIVE, Bethesda, MD 20892, USA
FEATURES
             Location/Qualifiers
                 source          1..930
                                 /organism="Homo sapiens"
                                 /db_xref="taxon:9606"
                                 /chromosome="12"
                                 /map="12p13"
                                 .1..21
primer_bind primer_1            1..930
CDS           1..930
              /note="G protein-coupled receptor; PCR-derived sequence; similar to an interval in BMC AC006518"
              /codon_start=1
              /product="candidate taste receptor T2R8"
              /protein_id="AAF43907.1"
              /db_xref="GI:7262615"
```

1. **Introduction**  
 2. **Background**  
 3. **Methodology**  
 4. **Results**  
 5. **Discussion**  
 6. **Conclusion**  
 7. **References**  
 8. **Appendix**  
 9. **Figure 1**  
 10. **Figure 2**  
 11. **Figure 3**  
 12. **Figure 4**  
 13. **Figure 5**  
 14. **Figure 6**  
 15. **Figure 7**  
 16. **Figure 8**  
 17. **Figure 9**  
 18. **Figure 10**  
 19. **Figure 11**  
 20. **Figure 12**  
 21. **Figure 13**  
 22. **Figure 14**  
 23. **Figure 15**  
 24. **Figure 16**  
 25. **Figure 17**  
 26. **Figure 18**  
 27. **Figure 19**  
 28. **Figure 20**  
 29. **Figure 21**  
 30. **Figure 22**  
 31. **Figure 23**  
 32. **Figure 24**  
 33. **Figure 25**  
 34. **Figure 26**  
 35. **Figure 27**  
 36. **Figure 28**  
 37. **Figure 29**  
 38. **Figure 30**  
 39. **Figure 31**  
 40. **Figure 32**  
 41. **Figure 33**  
 42. **Figure 34**  
 43. **Figure 35**  
 44. **Figure 36**  
 45. **Figure 37**  
 46. **Figure 38**  
 47. **Figure 39**  
 48. **Figure 40**  
 49. **Figure 41**  
 50. **Figure 42**  
 51. **Figure 43**  
 52. **Figure 44**  
 53. **Figure 45**  
 54. **Figure 46**  
 55. **Figure 47**  
 56. **Figure 48**  
 57. **Figure 49**  
 58. **Figure 50**  
 59. **Figure 51**  
 60. **Figure 52**  
 61. **Figure 53**  
 62. **Figure 54**  
 63. **Figure 55**  
 64. **Figure 56**  
 65. **Figure 57**  
 66. **Figure 58**  
 67. **Figure 59**  
 68. **Figure 60**  
 69. **Figure 61**  
 70. **Figure 62**  
 71. **Figure 63**  
 72. **Figure 64**  
 73. **Figure 65**  
 74. **Figure 66**  
 75. **Figure 67**  
 76. **Figure 68**  
 77. **Figure 69**  
 78. **Figure 70**  
 79. **Figure 71**  
 80. **Figure 72**  
 81. **Figure 73**  
 82. **Figure 74**  
 83. **Figure 75**  
 84. **Figure 76**  
 85. **Figure 77**  
 86. **Figure 78**  
 87. **Figure 79**  
 88. **Figure 80**  
 89. **Figure 81**  
 90. **Figure 82**  
 91. **Figure 83**  
 92. **Figure 84**  
 93. **Figure 85**  
 94. **Figure 86**  
 95. **Figure 87**  
 96. **Figure 88**  
 97. **Figure 89**  
 98. **Figure 90**  
 99. **Figure 91**  
 100. **Figure 92**  
 101. **Figure 93**  
 102. **Figure 94**  
 103. **Figure 95**  
 104. **Figure 96**  
 105. **Figure 97**  
 106. **Figure 98**  
 107. **Figure 99**  
 108. **Figure 100**  
 109. **Figure 101**  
 110. **Figure 102**  
 111. **Figure 103**  
 112. **Figure 104**  
 113. **Figure 105**  
 114. **Figure 106**  
 115. **Figure 107**  
 116. **Figure 108**  
 117. **Figure 109**  
 118. **Figure 110**  
 119. **Figure 111**  
 120. **Figure 112**  
 121. **Figure 113**  
 122. **Figure 114**  
 123. **Figure 115**  
 124. **Figure 116**  
 125. **Figure 117**  
 126. **Figure 118**  
 127. **Figure 119**  
 128. **Figure 120**  
 129. **Figure 121**  
 130. **Figure 122**  
 131. **Figure 123**  
 132. **Figure 124**  
 133. **Figure 125**  
 134. **Figure 126**  
 135. **Figure 127**  
 136. **Figure 128**  
 137. **Figure 129**  
 138. **Figure 130**  
 139. **Figure 131**  
 140. **Figure 132**  
 141. **Figure 133**  
 142. **Figure 134**  
 143. **Figure 135**  
 144. **Figure 136**  
 145. **Figure 137**  
 146. **Figure 138**  
 147. **Figure 139**  
 148. **Figure 140**  
 149. **Figure 141**  
 150. **Figure 142**  
 151. **Figure 143**  
 152. **Figure 144**  
 153. **Figure 145**  
 154. **Figure 146**  
 155. **Figure 147**  
 156. **Figure 148**  
 157. **Figure 149**  
 158. **Figure 150**  
 159. **Figure 151**  
 160. **Figure 152**  
 161. **Figure 153**  
 162. **Figure 154**  
 163. **Figure 155**  
 164. **Figure 156**  
 165. **Figure 157**  
 166. **Figure 158**  
 167. **Figure 159**  
 168. **Figure 160**  
 169. **Figure 161**  
 170. **Figure 162**  
 171. **Figure 163**  
 172. **Figure 164**  
 173. **Figure 165**  
 174. **Figure 166**  
 175. **Figure 167**  
 176. **Figure 168**  
 177. **Figure 169**  
 178. **Figure 170**  
 179. **Figure 171**  
 180. **Figure 172**  
 181. **Figure 173**  
 182. **Figure 174**  
 183. **Figure 175**  
 184. **Figure 176**  
 185. **Figure 177**  
 186. **Figure 178**  
 187. **Figure 179**  
 188. **Figure 180**  
 189. **Figure 181**  
 190. **Figure 182**  
 191. **Figure 183**  
 192. **Figure 184**  
 193. **Figure 185**  
 194. **Figure 186**  
 195. **Figure 187**  
 196. **Figure 188**  
 197. **Figure 189**  
 198. **Figure 190**  
 199. **Figure 191**  
 200. **Figure 192**  
 201. **Figure 193**  
 202. **Figure 194**  
 203. **Figure 195**  
 204. **Figure 196**  
 205. **Figure 197**  
 206. **Figure 198**  
 207. **Figure 199**  
 208. **Figure 200**  
 209. **Figure 201**  
 210. **Figure 202**  
 211. **Figure 203**  
 212. **Figure 204**  
 213. **Figure 205**  
 214. **Figure 206**  
 215. **Figure 207**  
 216. **Figure 208**  
 217. **Figure 209**

KAKRKTNLITWSCRVNKTQHASTKLFENLATLIPFCVCLMSFFLITLSLRHIRMQL  
ATGCRDPSTEAHVRAKAVISFLLLFIAYIYLSFLIATSSYEMETELAVIFGESIAL





Consensus quality: 145651 bases at least Q30  
 Consensus quality: 151922 bases at least Q20  
 Estimated insert size: 154982; sum-of-contigs estimation  
 Estimated insert size: 135000; agarose-fp estimation  
 Quality coverage: 3.9x in Q20 bases; agarose-fp estimation  
 Quality coverage: 3.4x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a working draft sequence. It currently  
 \* consists of 20 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 28948: contig of 28948 bp in length
* 28949 29048: gap of unknown length
* 29049 47981: contig of 18933 bp in length
* 47982 48081: gap of unknown length
* 48082 65432: contig of 17351 bp in length
* 65433 65532: gap of unknown length
* 65533 75793: contig of 10261 bp in length
* 75794 75893: gap of unknown length
* 75894 86770: contig of 10877 bp in length
* 86771 86870: gap of unknown length
* 86871 96590: contig of 9720 bp in length
* 96591 96690: gap of unknown length
* 96691 105444: contig of 8754 bp in length
* 105445 105544: gap of unknown length
* 105545 110791: contig of 5247 bp in length
* 110792 110891: gap of unknown length
* 110892 121864: contig of 10973 bp in length
* 121865 121966: gap of unknown length
* 121966 130966: contig of 9002 bp in length
* 130967 131067: gap of unknown length
* 131067 137246: contig of 6080 bp in length
* 137247 137247: gap of unknown length
* 137247 141648: contig of 4402 bp in length
* 141649 141748: gap of unknown length
* 141749 145148: contig of 3400 bp in length
* 145149 145248: gap of unknown length
* 145249 149349: contig of 4101 bp in length
* 149350 149449: gap of unknown length
* 149450 152521: contig of 3072 bp in length
* 152522 152621: gap of unknown length
* 152622 155479: contig of 2858 bp in length
* 155480 155579: gap of unknown length
* 155580 157339: contig of 1760 bp in length
* 157340 157439: gap of unknown length
* 157440 159496: contig of 2057 bp in length
* 159497 159596: gap of unknown length
* 159597 160694: contig of 1098 bp in length
* 160695 160794: gap of unknown length
* 160795 162553: contig of 1759 bp in length.

```

FEATURES  
 source  
 1. 162553  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="12"

BASE COUNT 50057 a 30129 c 29217 g 51223 t 1927 others  
 ORIGIN

alignment\_scores:  
 Quality: 376.00 Length: 294  
 Ratio: 1.979 Gaps: 5  
 Percent Similarity: 64.626 Percent Identity: 32.653

alignment\_block:  
 US-09-510-332-1 x AC016145/rev ..

Align seg 1/1 to reverse of: AC016145 from: 1 to: 162553

```

12 Leu1Val1IleGlnPheLeuLeuGlyIlePheThrAsnGlyIleIleVal 28
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58417 ATGTGTAATCGCATGACGCTGCTGCAAGAAAGAGAAATGCTCCCTTG 58368
58467 TTAGCAGTGGAGAGATTTTCAGTGGGATCTTACGAAATGATTCATTTGG 58418
28 IValValAsnGlyIleAspLeuIleLeuIleValIleValIleValIleVal 45
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58417 ATGTGTAATCGCATGACGCTGCTGCAAGAAAGAGAAATGCTCCCTTG 58368
45 sPheLeuLeuSerCysLeuAlaValSerArgIlePheLeuGlnLeuPhe 61
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58367 ATTTAATCTTCACAGCTGCTGCAATATTCAGAAATTTCTCATTTGCGCTA 58318
62 IlePheValAlaAsnValIleValIlePhePheIleGluPheIleMetCys 78
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58317 ATACTAATTAGATTGTTTATATTGCTGCTATATCCAGATGCTATGCTCAC 58268
78 sSerAlaAsnCysAlaIleLeuLeuPheIle.....AsnGlu 91
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58267 TGTAAAGAAATGAGAAATCATTTGACTCTTCTGAGACATTAACCAATCAT 58218
91 eugLeuLeuTrpLeuAlaThrTrpLeuGlyValPheTyrCysAlaIleVal 107
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58217 TAATATCTGTTGTCACACCTGCTCAGCATTTACTATTTCTTCTTCAAGATA 58168
108 AlaSerValArgHisProLeuPheIleTrpLeuIleValSerArgIleSer 124
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58167 GGTAAATTTCTTTCACCCATTTTCCTGATGAGAGGAGAAATTTGACAG 58118
124 sLeuValProTrpMetIleLeuGlySerLeuLeuTyrValSerMetIleC 141
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58117 GGTATTTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTTCT 58080
141 yValPheHisSerLysTyrAla.....Gly 149
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58079 CTGCTTTATTTAGCTTCCAGCAGCAGTGAATTTGAACGCTGATTTCAAG 58030
150 PheMetValProTyrPheLeuArgLysPhePheSerGlnAsnAlaThrI 166
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58029 TTTTGTGTGAAGCAAGAGAGAAACAACTTAATTTGAGAGAGAGAGAGAG 57980
166 eGlnLysGluAspThrLeuAlaIleGlnIlePheSerPheValAlaGluP 183
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57979 AATTAATCAACATGATGCTTCAACCAATTTTCTCAACCTGCAACGC 57930
183 heSerValProLeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePhe 199
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57929 TG...CTCCCTTTTGTGTGCTGCTAATGCTTTTCTCTTGAATCTTC 57883
200 SerLeuGlyArgHisThrArgIleMetArgAsnThrValAlaGlySer 216
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57882 TCCCTCGGAGACATATCAGGCAATGCAAGCTCAGCCAGGAGGAGAG 57833
216 gValProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSer 233
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57832 AGACCCACACACAGAGAGCCCATGTGAGCCCTGAAGGCTGATTTCTCT 57783
233 heLeuIleLeuTyrPheSerHisCysMetIleLysValPheLeuSer 249
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57782 TCCCTTCTCTTATTTATGCTTATTTGCTTCTTCTTCTTCTTCTTCTTCT 57733
250 LeuLysPheHisIleArgPheIlePheLeuPhePheIleLeuValI 266
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57732 ACCTACTTATGCGACAGACGGAATTAGCTGTGATTTTGTGATGCTCAT 57683
266 eGly...IleTyrProSerGlyHisSerIleLeuLeuLeuGlyAsn 282
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57682 AGCTTAATCTACCCCTCAAGCTATTTATTTCTTAATCTAGGAGAA 57633
282 rOlLysLeuLysGlnAsnAlaLysPheLeu 292
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```



```

repeat_region /rpt_family="AT_rich"
14232..14649
/rpt_family="LIM4"
repeat_region /rpt_family="LIM4"
14652..14742
/rpt_family="(TA)n"
repeat_region 15287..15421
/rpt_family="L2"
repeat_region 16053..16237
/rpt_family="LIMB5"
repeat_region 16238..16272
/rpt_family="U2"
repeat_region 16273..17683
/rpt_family="LIMB5"
repeat_region 18072..18210
/rpt_family="LIM4A"
repeat_region 18679..18844
/rpt_family="LIM4"
repeat_region 18895..19193
/rpt_family="LIPAL6"
repeat_region /rpt_family="AT_rich"
19194..19332
/rpt_family="AT_rich"
repeat_region /rpt_family="LIPB2"
19658..20023
/rpt_family="LIM4A"
repeat_region /rpt_family="LIM4A"
20437..20931
/rpt_family="AT_rich"
repeat_region /rpt_family="LIM4"
2549..26960
/rpt_family="LIM4"
repeat_region 27024..27268
/rpt_family="LIM4"
repeat_region 27269..27511
/rpt_family="MER46"
repeat_region 27793..28087
/rpt_family="AluJo"
repeat_region /rpt_family="(GA)n"
28123..28124
/rpt_family="(TA)n"
repeat_region 28277..28381
/rpt_family="LIM4"
repeat_region 28368..28912
/rpt_family="LIM4"
repeat_region 28925..29052
/rpt_family="MULTA1"
repeat_region 29201..29267
/rpt_family="Alus"
repeat_region /rpt_family="(TAA)n"
29508..29559
/rpt_family="AluY"
repeat_region /rpt_family="AluY"
30369..30396
/rpt_family="AT_rich"
repeat_region 31146..31316
/rpt_family="LIP2"
repeat_region 31857..32070
/rpt_family="AluJo"
repeat_region /rpt_family="LIM4"
32384..33480
/rpt_family="LIM4"

```

```

alignment_scores:
Quality: 376.00 Length: 320
Ratio: 2.000 Gaps: 10
Percent Similarity: 58.750 Percent Identity: 31.562

```

Alignment block:  
US-09-510-332-1 x AC006518/rev ..

Align seg 1/1 to reverse of: AC006518 from: 1 to: 173735

```

8 11eTyPheLeuLeuAlaValrIlegInPheLeuLeuGlyIlePheThrAs 24
|||||:|||||:|:|:|:|||||:|||||:|:|:|:|
5870 ATCTTTATTAATCTTAATAGTGAAGATTCTACTAGCAATATTTGGGAA 5821

```

```

24 nGlyIleIleValAlaValnGlyIleAsPheLeuIleLysArgLysM 41
|||||:|||||:|:|:|:|||||:|||||:|:|:|:|
5820 TGGATATCATCTGACTAGTCACTGATGAGTGAATTAAGAAAAAGA 5771
41 eTAlaPProLeuAsPheLeuLeuSerCysLeuAlaValSerArgIlePhe 57
|||||:|||||:|:|:|:|||||:|||||:|:|:|:|
5770 TTTCACAGTTGACTACATCTTACCAATTTAGTTATGCCCAAAATTTGT 5721
58 LeuGlnLeuPheIlePheTyGValAsnValIleValIlePheIleG 74
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5720 TTG...ATCAGTGTAAAGTGTGAATAGCATGTGAATAGTGAACCC 5674
74 uPheIleMerCysSerAlaAsnCysAlaIleLeuLeu..... 86
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5673 AGATGTTTATATACAAAATAACACAGATATCATTTTACCTTGCGA 5624
87 ..PheIleAsnGlnLeuGlnLeuTriPheLeuAlaThrTrpLeuGlyValPhe 102
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5623 CATTTGCCAATCTTAAATATGTGGATTACACCTGCTTAATGTCTTC 5574
103 TyGValAlaLysValAlaSerValArgHisProLeuPheIleTrpLeuTy 119
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5573 TATTTCTGAAATAGCAGATTCCTCTCATCCACTTTTCTCTGCTGTA 5524
119 sMetArgIleSerLysLeuValProTrpMetIleLeuGlySerLeu... 134
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5523 GTGGAAAATTTGATGTGTGCTGATCCCTGCGGATGCTTTGGCA 5474
135 ....LeuTyGValSerMetIleCysVal..... 142
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5473 TTTCCTGTGTTGGTACCTTATAGCAGCAATAGTACTGAGTTGTGATTT 5424
143 ...PheHisSer.....LysTyGAl 148
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5423 AGCTTTGACATGATGCGCAACATAAACAAATTAAGTAATTTGCA 5374
148 aGlyPheMetValProTyPheLeuArgLysPheSerGlnAsnAlaTr 165
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5373 TGTGAGTAAATATCCACTTT..... 5352
165 hTrIleGlnLysGlnAspThrLeuAlaIleGlnIlePheSerPheValAla 181
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5351 .....GACCCCTTGAATCTTAACTGTTGCAAT..... 5319
182 GluPheSerValProLeuLeuIlePheLeuPheAlaValLeuLeuI 198
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5318 .....GTCCCAATTAATGTGTGCTGATGATCATTTTCTTTAGT 5278
198 ePheSerLeuGlyArgHisThrArgGlnMetArgAsnThrValAlaGly 215
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5277 AAGATCTTTATGAGACATACCAACAAATAAATCTATGCTACGGCA 5228
215 eArgValProGlyArgGlyAlaProIleSerAlaLeuLeuSerIleLeu 231
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5227 GTAGGACCCACACACAGATGATGAGACCATTAATAAATGACT 5178
232 SerPheLeuIleLeuTyPheSerHisCysMetIleLysValPheLeu 248
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5177 TCATTTATCTCTTTTCTTTCTTATAC.....TATATTTTC 5143
248 rSerLeuLysPheHisIleArgArgPheIlePheLeu..... 260
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5142 TTCTAAT.....TTGATGACCTTTAGCTATCTATGACAAATATACA 5102
261 .....PhePheIleLeuValIleGlyIleTyPheProSerGly 272
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5101 AGTACGCTGTGACTTTGGAGAGATTCACGAATTCCTACCCCTGGGT 5052
273 HisSerLeuIleLeuIleLeuGlyAsnProLysLeuLysGlnAsnAlaTy 289
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5051 CACTACCTTATTTAATTTGTTTAAATATTAATTAACGTGAGCAGACATTTGT 5002

```

```

289 slyspheleu 292
      :|||||
5001 CAGATGCTG 4992

seq_name: gb_pr4:AF227135
seq_documentation_block:
LOCUS   AF227135      939 bp      DNA      PRI      18-MAR-2000
DEFINITION Homo sapiens candidate taste receptor T2R9 gene, complete cds.
ACCESSION AF227135
VERSION   AF227135.1  GI:7262616
KEYWORDS
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 939)
  Adler, E., Hoon, M.A., Mueller, K.L., Chandrasekar, J., Ryba, N.J. and
  Zuker, C.S.
  A novel family of mammalian taste receptors
  Cell 100 (6), 693-702 (2000)
  2 (bases 1 to 939)
  Adler, E., Hoon, M.A., Mueller, K.L., Chandrasekar, J., Ryba, N.J. P.
  and Zuker, C.S.
  Direct Submission
  Submitted (21-JAN-2000) NIDCR, NIH, Bldg. 10 Room 1N106, 10 Center
  Drive, Bethesda, MD 20892, USA
FEATURES
  Source
    1..939
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="12"
    /map="12p13"
    1..19
    1..939
    /note="G protein-coupled receptor; PCR-derived sequence;
    similar to an interval in BAC AC006518"
    /codon_start=1
    /product="candidate taste receptor T2R9"
    /protein_id="AAF43908.1"
    /db_xref="GI:7262617"
    /translation="MPSAIEAIVIIILAGELTIGMGFIVLVNCIDMLKRRDISL
    DIIISLAISRICLCVSLDGFMLLEPRGTGNSVIVNVTFFANNSLMETSC
    LSTFLKIANISHEFEFWLKLKNKWMALILSFLISLISPKNDMMYHLEKVS
    HEENTWKEKVKSLPGTEKQTLNGVAVPFLICISFFLLSLVHTKQIRLHATG
    FRDPTSEAHMRAIKAVIIFLLILVYVPLVMTSSALIPQKLVLMIGDIVTIFPS
    SHSFLINGSKLREAFKMLRFVCFLLRRKPEVP"
    <1..>939
    /product="candidate taste receptor T2R9"
  primer_bind
    BASE COUNT 248 a 187 c 186 g 318 t
    ORIGIN
      mRNA
      /product="candidate taste receptor T2R9"
      919..939
      primer_bind
        248 a 187 c 186 g 318 t
      alignment_scores:
        Quality: 371.50      Length: 314
        Ratio: 1.905      Gaps: 9
        Percent Similarity: 62.102      Percent Identity: 32.166
      alignment_block:
        US-09-510-332-1 x AF227135
      Align seg 1/1 to: AF227135 from: 1 to: 939
      8 1leTyRpheLeuAlaValIleGlnPheLeuLeuGlyIlePheThAs 24
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      22 ATATATATATATTTAATGCGTGGATGACCAATGAGGATTTGGGAAA 71
      24 nglytlelleValValValasnglylleaspleulleyshtsargylsm 41
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      72 TGGATTCATCTAGTACTGATGCTGCTCAAAAGAGAGATA 121

```

```

41 etAlaProLeuAspLeuLeuLeuSerCysLeuAlaValSerArgIlePhe 57
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
122 TTTCCTGATGATGACATCATCTGATGACCTTGCCCATCTCCAGATCTGT 171
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
58 LeuGlnLeuPhele.....PheTyRValAsnValIleValI 70
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
172 CTGCTGTGTGTAATATATAGATGCGCTCTTATGCTGCTCTTCCAGG 221
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
70 ePhePheIleGluPheIleMetCysSerAlaAsnCysAlaIleLeuLeuP 87
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
222 TACATATGGCAATAGCGCTAGTAAAGCATTTGCAATGTTGTGCGCAT 271
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
87 heIleAsnGluLeuGluLeuTyRLeuAlaThrTyRLeuGlyValPheTyR 103
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
272 TTGCCATATATTCAGTCTCTGCTTACTTCTTCCGACATCTTCTAT 321
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104 CysAlaValAlaSerValArgHisProLeuPheIleTyRLeuPheLysme 120
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
322 TTACTCAAGATAGCCAAATATATCGCACCCATTTTCTTGGCTGAAGCT 371
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
120 tArgIleSerLysLeuValProTyrPheIleLeuGlySerLeuLeuTyRv 137
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
372 AAAGATCAACAGGTCATGCTTGGCATTTCTTGCGGCTCTTCTT...A 418
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
137 alSerMetIleCysValPheHisSerLysTyRAlaGlyPheMetValPro 153
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
419 TCTCTTTAATT.....ATTAGTGTTCGA 441
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
154 .....TyrPheLeuArgLysPhe...PheSerGlnAs 163
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
442 AAGATGATGATATGTGTGATCATCCTTTCAAAAGTCAGTCATGAAGAAB 491
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
163 nAlaThrIleGln.....LysGlnA 170
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492 CATTCCTTGGAATTCAAAGTCAGTAAATTCAGCTACTTCCAAACAGT 541
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
170 spThrLeuAlaIleGlnIlePheSerPheValAlaGluPheSerValPro 186
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
542 TAACCTGTAACCTGGGGGTGATG.....GTTCC 570
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
187 LeuLeuIlePheLeuPheAlaValLeuLeuIlePheSerLeuGlyAr 203
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
571 TTTATCCTTGGCTGATCTCATTTTCTGTGATCTTTCTCCCTAGTTAG 620
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
203 gHisThrArgIleMetArgAsnThrValAlaGlySerArgValProGlyA 220
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
621 ACACACCAACAGCATTCGATGCTACAGGGTTCAGACACCCAGATA 670
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
220 rGlyValAProIleSerAlaLeuLeuSerIleLeuSerPheLeuIleLeu 236
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
671 CAGAGCCCAACATGAGGCCCATGAAGCAGTATCTTTCTGCTCCCTC 720
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
237 .....TyrPheSerHisCysMetIleLysValPheLeuSerSerIe 250
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
721 CTCATCGTGTACTACCA.....GTCTTCTGTTATGACCTCTAGCCG 764
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
250 slyPheHisIleArgArgPheIlePheLeuPhePheIleLeuValIleG 267
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
765 TCGATTCCTCCAGGAATAATAGTGTGATGATGCTGACATAGTAATC 814
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
267 LytleTyRProSerGlyHisSerLeuIleLeuIleLeuGlyAsnProLys 283
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
815 TCATTTTCCCATCAAGCCATTCATCTTCAATTAATGAGGAATAGCAAG 864
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
284 LeuLysGlnAsnAlaLysLysPheLeuLeuHisSerLysCys 297
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
865 TTGAGGAGAGCTTTTCTGAAAGATGTTAAGATTTGTGTAAGTGT 906

```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 12:57:24 : Search time 71 Seconds  
(without alignments)  
493.594 Million cell updates/sec

Title: US-09-510-332-1  
Perfect score: 1521  
Sequence: 1 MESHLLIYFLAVIQLG.....GNPKIKQAKKFLSKCCQ 299

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 segs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP-archaea:\*\*\*  
2: SP-bacteria:\*\*\*  
3: SP-fungi:\*\*\*  
4: SP-human:\*\*\*  
5: SP-invertebrate:\*\*\*  
6: SP-mammal:\*\*\*  
7: SP-mhc:\*\*\*  
8: SP-organelle:\*\*\*  
9: SP-phage:\*\*\*  
10: SP-plant:\*\*\*  
11: SP-rodent:\*\*\*  
12: SP-virus:\*\*\*  
13: SP-vertebrate:\*\*\*  
14: SP-unclassified:\*\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1521	100.0	299	4 Q9NYW7	Q9NYW7 homo sapien
2	762	51.4	335	11 Q9JKT2	Q9JKT2 mus musculu
3	767	50.4	335	11 Q9JKT1	Q9JKT1 rattus norv
4	405	26.6	316	4 Q9NYW6	Q9NYW6 homo sapien
5	376	24.7	309	4 Q9NYW2	Q9NYW2 homo sapien
6	376	24.7	318	4 Q9NYW3	Q9NYW3 homo sapien
7	371.5	24.4	312	4 Q9NYW1	Q9NYW1 homo sapien
8	369	24.3	297	11 Q9JKE9	Q9JKE9 rattus norv
9	358	23.5	307	4 Q9NYW0	Q9NYW0 homo sapien
10	356	23.4	303	4 Q9NTV9	Q9NTV9 homo sapien
11	355.5	22.3	308	11 Q9JKE7	Q9JKE7 rattus norv
12	339.5	22.3	314	11 Q9JKT6	Q9JKT6 rattus norv
13	331.5	20.8	305	11 Q9JKT7	Q9JKT7 rattus norv
14	317.5	20.9	308	11 Q9JKT9	Q9JKT9 rattus norv
15	317	20.8	309	11 Q9JKT5	Q9JKT5 rattus norv
16	315.5	20.7	299	4 Q9NYW4	Q9NYW4 homo sapien
17	315.5	20.7	333	11 Q9JKT0	Q9JKT0 rattus norv
18	313	20.6	291	4 Q9NTV7	Q9NTV7 homo sapien
19	310.5	20.4	317	4 Q9NYW8	Q9NYW8 homo sapien

20	298	19.6	299	4 Q9NYW5	Q9NYW5 homo sapien
21	294	19.3	300	11 Q9JKT4	Q9JKT4 mus musculu
22	289	19.0	309	11 Q9JKT8	Q9JKT8 rattus norv
23	266.5	17.5	299	11 Q9JKT0	Q9JKT0 rattus norv
24	264.5	17.4	297	11 Q9JKT3	Q9JKT3 mus musculu
25	242.5	15.9	246	11 Q9JKT0	Q9JKT0 mus musculu
26	237.5	15.6	245	11 Q9JKT1	Q9JKT1 mus musculu
27	231.5	15.2	245	11 Q9JKT4	Q9JKT4 mus musculu
28	230.5	15.2	243	11 Q9JKT2	Q9JKT2 mus musculu
29	217.5	14.3	238	11 Q9JKT8	Q9JKT8 mus musculu
30	182	12.0	131	11 Q9JKE8	Q9JKE8 rattus norv
31	134	8.8	410	13 Q9PMU2	Q9PMU2 typhloecte
32	130.5	8.6	310	11 Q9PMU3	Q9PMU3 mus musculu
33	128.5	8.4	354	13 Q9PT44	Q9PT44 brachydanto
34	125	8.2	287	13 Q9PT44	Q9PT44 pantodon bu
35	125	8.2	352	13 Q9JKT5	Q9JKT5 labetrophe
36	125	8.2	382	13 Q9JKT8	Q9JKT8 caenorhabdi
37	124	8.2	346	13 Q9PUA9	Q9PUA9 bufo japoni
38	124	8.2	500	8 Q9ZQZ2	Q9ZQZ2 cyanidlosch
39	122.5	8.1	335	11 Q6Z853	Q6Z853 rattus norv
40	122.5	8.1	547	5 Q45965	Q45965 caenorhabdi
41	122	8.0	352	13 Q9JKT5	Q9JKT5 dimidiocbro
42	121.5	8.0	327	11 Q70271	Q70271 rattus norv
43	121.5	8.0	341	8 Q9MD83	Q9MD83 dirosophila
44	120.5	7.9	327	11 Q70269	Q70269 rattus norv
45	120.5	7.9	342	12 Q90387	Q90387 kaposi's sa

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	299 AA.
Q9NYW7	Q9NYW7			
AC	Q9NYW7			
DT	01-OCR-2000 (TREMBLrel. 15, Created)			
DT	01-OCR-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCR-2000 (TREMBLrel. 15, Last annotation update)			
DE	CANDIDATE TASTE RECEPTOR T2RL			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Adler E., Hoon M.A., Mueller K.L., Chandrasekhar J., Ryba N.J.P.,			
RA	Zuker C.S.,			
RT	"A novel family of mammalian taste receptors."			
RL	Cell 100:693-702(2000).			
DR	EMBL: AF227129; AAF43902.1; -			
KW	Receptor.			
SEQ	SEQUENCE 299 AA; 34333 MW; CFB17DD15645392C CRC64;			

Query Match	Score 1521:	DB 4:	Length 299:
Best Local Similarity	100.0%	Pred. No. 2.5e-97;	
Matches 299; Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1 MESHLLIYFLAVIQLGIFNGIIYVNGIDIKRKNAPDLLSLCLAVSRIFQL 60		
DB	1 MESHLLIYFLAVIQLGIFNGIIYVNGIDIKRKNAPDLLSLCLAVSRIFQL 60		
QY	61 FIFYVNVIVIFIEFIMCSANCAILLFETNELMLATWLGVYCAKVASVRRPFTWKM 120		
DB	61 FIFYVNVIVIFIEFIMCSANCAILLFETNELMLATWLGVYCAKVASVRRPFTWKM 120		
QY	121 RISKLVPMMLISLIVSMICVHSKYAGFVYFLRKFSSQATTOEDTLAIOFSEV 180		
DB	121 RISKLVPMMLISLIVSMICVHSKYAGFVYFLRKFSSQATTOEDTLAIOFSEV 180		
QY	181 AERSVPLILFLFAVLLIFSLGHRTRQMRNTVAGSRVPGRGAPISALLSIFLLIFSH 240		
DB	181 AERSVPLILFLFAVLLIFSLGHRTRQMRNTVAGSRVPGRGAPISALLSIFLLIFSH 240		

OY 241 CMIKVFLSLKFIHRRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCCQ 299  
 DB 241 CMIKVFLSLKFIHRRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCCQ 299

RESULT 2  
 OYJKT2 PRELIMINARY; PRT; 335 AA.  
 AC 09JKT2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R19.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SVJ;  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrasekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227149; AAF43922.1; -  
 KW Receptor  
 SQ SEQUENCE 335 AA; 37734 MW; 1399C323FC7CE645 CRC64;

Query Match 51.4%; Score 782; DB 11; Length 335;  
 Best Local Similarity 51.5%; Pred. No. 1.2e-46;  
 Matches 154; Conservative 51; Mismatches 90; Indels 4; Gaps 3;

OY 1 MESHLLIYFLAVIQFLGFTNGIIVVNGIDLIKRRMAPDLISCLAVSRIFQL 60  
 DB 1 MMEGHMFLFLVAVQFLGVLNGLIIVVNAIDLIKMKKAPDLISCLAVSRIFQL 60  
 OY 61 FIFVYVYVIFEFIFMCSANCAILFELNELMTWLCVFCAYASVRHPLFTWLM 120  
 DB 61 CILFAQGLSFLVAVQFLGVLNGLIIVVNAIDLIKMKKAPDLISCLAVSRIFQL 120  
 OY 121 RISKLVPMILGSLVSMICVFSKYGAFVYFLKRFSSQNAI-TQKEDTLAIQIFSF 179  
 DB 121 RISKLVPMILGSLVSMICVFSKYGAFVYFLKRFSSQNAI-TQKEDTLAIQIFSF 179  
 OY 121 RISKLVPMILGSLVSMICVFSKYGAFVYFLKRFSSQNAI-TQKEDTLAIQIFSF 179  
 DB 121 RISKLVPMILGSLVSMICVFSKYGAFVYFLKRFSSQNAI-TQKEDTLAIQIFSF 179  
 OY 180 VAESVPLIFLFAVLLIFSLGRHTRONRVAGSRVGRCAPISALISFLIYFS 239  
 DB 180 VAESVPLIFLFAVLLIFSLGRHTRONRVAGSRVGRCAPISALISFLIYFS 239  
 OY 181 --GUTLPLFLFAVLLIFSLGRHTRONRVAGSRVGRCAPISALISFLIYFS 237  
 DB 181 --GUTLPLFLFAVLLIFSLGRHTRONRVAGSRVGRCAPISALISFLIYFS 237  
 OY 240 HCMIKVFLSLKFIHRRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCC 298  
 DB 240 HCMIKVFLSLKFIHRRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCC 298  
 DB 238 HDMAVAVLITGQVLYLGSRTFVCLVIGMYPISLHIVLILGNPKIKONAKKFLHSHKCC 296

RESULT 3  
 OYJKT2 PRELIMINARY; PRT; 335 AA.  
 AC 09JKT2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Adler E., Hoon M.A., Mueller K.L., Chandrasekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).

DR EMBL: AF227140; AAF43913.1; -  
 KW Receptor  
 SQ SEQUENCE 335 AA; 37851 MW; AB8A1CEDDE70C12C CRC64;

Query Match 50.4%; Score 767; DB 11; Length 335;  
 Best Local Similarity 51.8%; Pred. No. 1.3e-45;  
 Matches 155; Conservative 53; Mismatches 87; Indels 4; Gaps 4;

OY 1 MESHLLIYFLAVIQFLGFTNGIIVVNGIDLIKRRMAPDLISCLAVSRIFQL 60  
 DB 1 MMEGHMFLFLVAVQFLGVLNGLIIVVNAIDLIKMKKAPDLISCLAVSRIFQL 60  
 OY 61 FIFVYVYVIFEFIFMCSANCAILFELNELMTWLCVFCAYASVRHPLFTWLM 120  
 DB 61 CILFAQGLSFLVAVQFLGVLNGLIIVVNAIDLIKMKKAPDLISCLAVSRIFQL 120  
 OY 121 RISKLVPMILGSLVSMICVFSKYGAFVYFLKRFSSQNAI-TQKEDTLAIQIFSF 179  
 DB 121 RISKLVPMILGSLVSMICVFSKYGAFVYFLKRFSSQNAI-TQKEDTLAIQIFSF 179  
 OY 121 RISKLVPMILGSLVSMICVFSKYGAFVYFLKRFSSQNAI-TQKEDTLAIQIFSF 179  
 DB 121 RISKLVPMILGSLVSMICVFSKYGAFVYFLKRFSSQNAI-TQKEDTLAIQIFSF 179  
 OY 180 VAESVPLIFLFAVLLIFSLGRHTRONRVAGSRVGRCAPISALISFLIYFS 239  
 DB 180 VAESVPLIFLFAVLLIFSLGRHTRONRVAGSRVGRCAPISALISFLIYFS 239  
 OY 179 VGLTLPFLFAVLLIFSLGRHTRONRVAGSRVGRCAPISALISFLIYFS 237  
 DB 179 VGLTLPFLFAVLLIFSLGRHTRONRVAGSRVGRCAPISALISFLIYFS 237  
 OY 240 HCMIKVFLSLKFIHRRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCC 298  
 DB 240 HCMIKVFLSLKFIHRRFIFLFIYIGIYPSGSHLILGNPKIKONAKKFLHSHKCC 298  
 DB 238 HDMAVAVLITGQVLYLGSRTFVCLVIGMYPISLHIVLILGNPKIKONAKKFLHSHKCC 296

RESULT 4  
 OYJKT2 PRELIMINARY; PRT; 316 AA.  
 AC 09JKT2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Adler E., Hoon M.A., Mueller K.L., Chandrasekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227130; AAF43903.1; -  
 KW Receptor  
 SQ SEQUENCE 316 AA; 35914 MW; 6A7BA3C6497ABC5B CRC64;

Query Match 26.6%; Score 405; DB 4; Length 316;  
 Best Local Similarity 36.4%; Pred. No. 8.1e-21;  
 Matches 114; Conservative 43; Mismatches 114; Indels 42; Gaps 8;

OY 8 IYFLAVIQFLGFTNGIIVVNGIDLIKRRMAPDLISCLAVSRIFQLFIYVNV 67  
 DB 8 IYFLAVIQFLGFTNGIIVVNGIDLIKRRMAPDLISCLAVSRIFQLFIYVNV 67  
 OY 68 IYFIFEFIFMCSANCAILFELNELMTWLCVFCAYASVRHPLFTWLM 119  
 DB 68 IYFIFEFIFMCSANCAILFELNELMTWLCVFCAYASVRHPLFTWLM 119  
 OY 67 ---FLIESPPTHDSGITMOTIDVSWFTTNHLSLWLAACGLVCLKIASFSHPFTWLM 123  
 DB 67 ---FLIESPPTHDSGITMOTIDVSWFTTNHLSLWLAACGLVCLKIASFSHPFTWLM 123  
 OY 120 MRISKLVPMILGSLVSMICVFSKYGAFVYFLKRFSSQNAI-TQKEDTLAIQIFSF 168  
 DB 120 MRISKLVPMILGSLVSMICVFSKYGAFVYFLKRFSSQNAI-TQKEDTLAIQIFSF 168  
 OY 124 WRVSRVYVWMLLGLALLSCGSTASLINEFKLYSVFGRGIEATRVNTEHPRKRSSVYLIHV 183  
 DB 124 WRVSRVYVWMLLGLALLSCGSTASLINEFKLYSVFGRGIEATRVNTEHPRKRSSVYLIHV 183  
 OY 169 EDTLAIQIFSVAVESVPLIFLFAVLLIFSLGRHTRONRVAGSRVGRCAPISALIS 228  
 DB 169 EDTLAIQIFSVAVESVPLIFLFAVLLIFSLGRHTRONRVAGSRVGRCAPISALIS 228  
 OY 184 LGTL-----WLPPLIVLASYSLLIFSLGRHTRONRVAGSRVGRCAPISALIS 234  
 DB 184 LGTL-----WLPPLIVLASYSLLIFSLGRHTRONRVAGSRVGRCAPISALIS 234



QY 229 SILSEF---LITSEHCMIKVFSL--KFHRRIFLEFVIGVPSGSLILGNPKL 284  
 DB 235 ILLSEFFLEFLYFLAFIAGNGLPRTKAKMIGEVMT---FTPAGHSFILILGNSKL 291  
 QY 285 KONAKKFLHSCK 297  
 DB 292 KOT---FVVMRLC 301

## RESULT 5

Q9NWM2 PRELIMINARY: PRT: 309 AA.  
 AC Q9NWM2. 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxId=9606;  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrashekar J., Ryba N.J.P.,  
 RT Zuker C.S.;  
 RL "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL; AF227134; AAF43907.1; -  
 KW Receptor.  
 SQ SEQUENCE 309 AA; 35877 MW; 5E3D94B726A52413 CRC64;

Query Match 24.7%; Score 376; DB 4; Length 309;  
 Best Local Similarity 31.6%; Pred. No. 7.7e-19;  
 Matches 101; Conservative 56; Mismatches 101; Indels 62; Gaps 10;

QY 8 IYELLAVIOFLGIFNGIIVVNGIDILKHKMAPLDLLSCLANSRIFLOLFIVVNV 67  
 DB 8 IYILLIGEFLTIGNGIYALVNMIDMKKISTVDYILNMLVARICL-LSVAVNG 66  
 QY 68 IYVFFIFMCSANCAILL-----FINELEMLATWLGVCYAKVASVRHPFLTWKMR1 122  
 DB 67 IYVILNDVYTKKQOIVITFTFNTFANYLMMWITTCIANVYFLKIKSSSHPLFLMKW1 126  
 QY 123 SKVPMMLISL--LYVSMICV-----FHS-----KXAGEVVPFLKFRS 161  
 DB 127 DMVVMHLLGFAISLIVSLIAIVLSCDYRFHAIKHKRNITEMPHVSKIPF----- 180  
 QY 162 ONATIQEDTLAIOESFVAFESVPLILFEAVLLIFSGRHTROMRNTVAGSRVPRG 221  
 DB 181 -----EPLTLFNLFL-----VPFVLSLISFPLVRSIMRHTKQIKLYATGSRDPSTE 228  
 QY 222 APISALLSILSEFLYFCHMKVFLSLKFIHRRIFL-----FLLVIGIYPSG 272  
 DB 229 VHVRAIKMTSFIFFFLY-----YISSI---LMTFSYLMTKYKLAVEFGHAIILYPLG 280  
 QY 273 HSLILILGNPKLONAKKFL 292  
 DB 281 HSLILIVLNKLRQTFVRL 300

## RESULT 6

Q9NWM3 PRELIMINARY: PRT: 318 AA.  
 AC Q9NWM3. 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxId=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrashekar J., Ryba N.J.P.,  
 RT Zuker C.S.;  
 RL "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL; AF227133; AAF43906.1; -  
 KW Receptor.  
 SQ SEQUENCE 318 AA; 36549 MW; F187ADB2D8274B8A CRC64;

Query Match 24.7%; Score 376; DB 4; Length 318;  
 Best Local Similarity 32.7%; Pred. No. 7.9e-19;  
 Matches 96; Conservative 60; Mismatches 120; Indels 18; Gaps 5;

QY 12 LAVIOFLGIFNGIIVVNGIDILKHKMAPLDLLSCLANSRIFLOLFIVVNVYIF 71  
 DB 12 LAVGEFSGILGNAGFGLVNCMDVKKRKIASIDLITSLAISRLCLVILDCFTLV 71  
 QY 72 FIEFIMCSANCAILLF-----NELEMLATWLGVCYAKVASVRHPFLTWKMRISKLP 127  
 DB 72 YPDVYATGKEMRIIDFHWTLNHLSTWFATCLSTIYFEKGNFHFPLMMKMRIDRVS 131  
 QY 128 WILGSLIYVSMICVFHSHKXA-----GEVVPFLKRFSSONATIQEDTLAIOESF 179  
 DB 132 WILGCV-----VLSVFLSPATENLNADFRFCYAKKRTMLTSCRVNKHASTFLFN 187  
 QY 180 VAEFVPLILFEAVLLILSLGRHTROMRNTVAGSRVPGAPISALLSILSFLIYFS 239  
 DB 188 LATL-LPFCVCLMSFLLILSLRHRIRMOISATGCDPSTEAVHRLAKVIFSLLEFLA 246  
 QY 240 HCMIKVFLSSLKFIHRRIFLEFVIG-IYPSGSHLLILGNPKLONAKKFL 292  
 DB 247 YLSFLITSSYFMPPEFLAVIFGESTALITYSSSHSFLILGNKLRHASKLY 300

## RESULT 7

Q9NWM1 PRELIMINARY: PRT: 312 AA.  
 AC Q9NWM1. 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R9.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxId=9606;  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrashekar J., Ryba N.J.P.,  
 RT Zuker C.S.;  
 RL "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL; AF227135; AAF43908.1; -  
 KW Receptor.  
 SQ SEQUENCE 312 AA; 35611 MW; CCE66F1E31051F49 CRC64;

Query Match 24.4%; Score 371.5; DB 4; Length 312;  
 Best Local Similarity 32.1%; Pred. No. 1.6e-18;  
 Matches 100; Conservative 60; Mismatches 113; Indels 39; Gaps 8;

QY 8 IYELLAVIOFLGIFNGIIVVNGIDILKHKMAPLDLLSCLANSRIFLOLFIVVNVYIF 63  
 DB 8 IYILLIGEFLTIGNGIYALVNCIDMKRDISLIDILSLAISRLCLCVISLDGF 67  
 QY 64 YVNVYVFFIEFIMCSANCAILLFINELEMLATWLGVCYAKVASVRHPFLTWKMRIS 123  
 DB 68 FMLLPFGTGNVSLVYVNVWTFPANNSSLMPTSCSIFYLKIANISHPFFFWLTKIN 127  
 QY 124 KLVPMMLISLIVSMICVFHSHKXAGPMP-----YFLAKF-FSONATIQ----- 167

DB 128 KVMALILGSL-LSL-----ISVKNDDMMVHLFKVSHENITWPKFKSIPGT 177  
 QY 168 -KEDFLAIQIFSVAFESVPLILFAVLLIFSLGCRTRORNTVASSRPGCAPISA 226  
 DB 178 FQGLTINLGV-----VFLICISLFFLSLVRHTKQRLHATFROPSTEAHRA 230  
 QY 227 LLSISFLILYSHC-MRKVPLSLKFRIRFIFLFLVIGIYPSGSHLLILGNPKLK 285  
 DB 231 IKAVIIFLLILVYVPFLVMTSSALIPQGLVLMIGDIVIYVFPSSHSFLINGNSKLR 290  
 QY 286 QNAKFFLHSC 297  
 DB 291 EAFKMLRFVKC 302

RESULT 8  
 QJKE9  
 ID 09JKE9 PRELIMINARY: PRT: 297 AA.  
 AC 09JKE9  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE TASTE RECEPTOR RT2R6 (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrasekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF240766; AAF45304.1; -.  
 KW Receptor.  
 FT NON\_TER  
 SQ SEQUENCE 297 AA; 34107 MW; 873D0440C2863E25 CRC64;

Query Match 24.3%; Score 369; DB 11; Length 297;  
 Best Local Similarity 34.1%; Pred. No. 2, 2e-18;  
 Matches 104; Conservative 51; Mismatches 106; Indels 44; Gaps 9;

QY 18 LLSLFTNGIIVVNGIDILIKHRKMAPDLLSLCLAVSRIFLQIFVYVNIYFIEFIM 77  
 DB 3 LVGLGNAFIALVNFMMKMKRKTALDILSLMSRILCQCIILDCIILVOYDPYV 62  
 QY 78 CSANCAILFT---NELEMLATWLGVEYCAVASYRHPFLTWLKRISKLVPMMILGS 133  
 DB 63 RGMKRIIDFMTLNLNLSVWFATCLSIFFFKIANEFHPLFMKIRIDKILRFLAC 122  
 QY 134 LLYVMICVFHSKYAGFVNY-----FLRKFFSQ--NATQ-----KEDTL 172  
 DB 123 L--LILSLC-----FSLVTEMLADDFRCKYKTERINSTLRCKLNAGASVAVNN 172  
 QY 173 AIOQFSVAESVPLILFAVLLIFSLGRHTRORNTVASSRPGCAPISALLSILS 232  
 DB 173 LVMLF-----PFSVSLVSF-----LLILSLRHTROMQVNTGVNDSTAHAKATAVIS 224  
 QY 233 FLIYFSHCIMKIVFSLKFRIRFIFLFLVIGIYPSGSHLLILGNPKLONAKKF 291  
 DB 225 FLVLFYICLAFLANSSYMPSESLAVIWEGLIALIYPSHSFILLINSKLOKASVAV 284  
 QY 292 LHSK 296  
 DB 285 LCRVK 289

RESULT 9  
 ID 09NTWO PRELIMINARY: PRT: 307 AA.  
 AC 09NTWO  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.  
 NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrasekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227136; AAF43909.1; -.  
 KW Receptor.  
 SQ SEQUENCE 307 AA; 35365 MW; 058DD69A55AE71C6 CRC64;

Query Match 23.5%; Score 358; DB 4; Length 307;  
 Best Local Similarity 31.0%; Pred. No. 1, 3e-17;  
 Matches 96; Conservative 68; Mismatches 110; Indels 36; Gaps 12;

QY 8 IYFLAVIOPLGIFTNGIIVVNGIDILIKHRKMAPDLLSLCLAVSRIFLQIFVYV 67  
 DB 8 IFEFVVSSEVGLNGFVGLVNCIDAKN-KLSYIGITLGLAISRIFL-IWIIIDG 65  
 QY 68 IVFFIEFIMCSANCAILFI-----NELEMLATWLGVEYCAVASYRHPFLTWL 120  
 DB 66 FIOIFSPNIYASGN--LIIYISYFVWYQSSMWPATSLISYFELKINFSYFLMKS 123  
 QY 121 RISKLVPMMILGSLYVSMICVFSKYGAFVNYFLRKFFSONATQKEDTL-AIQFS- 178  
 DB 124 RRMVLPFMIY--FLILSLNF-----AYAIINDYK--KNDYVMDLNMKS 169  
 QY 179 --FVAFESVPL-LIFLFAV-----LLIFSLGRHTRORNTVASSRPGCAPISALLS 230  
 DB 170 EYFIQIILNLNGVIFEFLLSLTICIFLISLRHNRQMSVNTGIRDSTAHAVAKVYL 229  
 QY 231 LSLFLYFSHCIMKIVFSLKFRIRFIFLFLVIGIYPSGSHLLILGNPKLONAK 289  
 DB 230 ISFIFLFLYFGMAIEISCFVRENKLLIMFGMTTITVWGHSPILLINSKLOKASL 289  
 QY 290 KFLHSCQ 299  
 DB 290 RVLOQLKCE 299

RESULT 10  
 ID 09NTV9 PRELIMINARY: PRT: 303 AA.  
 AC 09NTV9  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE CANDIDATE TASTE RECEPTOR T2R13.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.  
 NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrasekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227137; AAF43910.1; -.  
 KW Receptor.  
 SQ SEQUENCE 303 AA; 35118 MW; 756BF5382E36BF5 CRC64;

Query Match 23.4%; Score 356; DB 4; Length 303;  
 Best Local Similarity 31.9%; Pred. No. 1, 8e-17;  
 Matches 97; Conservative 69; Mismatches 116; Indels 22; Gaps 11;



RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227144; AAF43917.1;  
 KW Receptor  
 SQ SEQUENCE 305 AA; 35594 MW; 533BABD8C8BA4BE9 CRC64;

Query Match 21.8%; Score 331.5; DB 11; Length 305;  
 Best Local Similarity 28.7%; Pred. No. 8,5e-15;  
 Matches 86; Conservative 76; Mismatches 123; Indels 13; Gaps 8;

QY 8 IYFLAVI---QFLGIFTNGIIVVNGIDILKHKMAPDLISCLAVSRIFL-QLEFI 62  
 DB 5 LYDLITVIMAEFIEFGNTNGFIYLTNCIAWLSKRTLSFICMIOQLFALISVVLWIMELL 64  
 QY 63 FYVAVIYFIEFIEFIMCSANCAILFI--NELEMLATWLGVCYAKVASVHPFLTKM 120  
 DB 65 AWLKYMKISFYSLAGTELRYMMLWVSNHPSLWLAITLSFYLLKIASFSRPVFLYK 124  
 QY 121 RISLKVPMMLIGSLLYSM-ICVFHSKYAGFMPVFLRKFSQNAITQKEDTLAIOFSE 179  
 DB 125 RVKVVLLILILGNLIFLNFILQINTHIEDMDQY--KRNTWMSRVNFEVGSNVLLE 182  
 QY 180 VAERSV-PLTFLFAVLLITSLGRHROMKNTVAGSRVGRGAPISALISLILYF 238  
 DB 183 MIMSVLPFTALVSFILLFLSLMKHLQKMHLSRGERDPTKHAVALRIMVSFLLYA 242  
 QY 239 SHCMIKVFLSLKFIHRFIFLFIYIGI-YPSGHSILILINPKRLKONAKKRLHSHK 297  
 DB 243 TY-FLSPFISLIAPMAHKGDMLMSLVGLFYPSSHFIILLGHSNLRHSSCLVITYLR 301

RESULT 14  
 Q9JKT9 PRELIMINARY; PRT; 308 AA.  
 AC Q9JKT9;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 OS CANDIDATE TASTE RECEPTOR T2R4.  
 OC Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrashekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227142; AAF43915.1;  
 KW Receptor.  
 SQ SEQUENCE 308 AA; 35052 MW; 00BC4193204E043 CRC64;

Query Match 20.9%; Score 317.5; DB 11; Length 308;  
 Best Local Similarity 29.4%; Pred. No. 7,8e-15;  
 Matches 89; Conservative 61; Mismatches 128; Indels 25; Gaps 9;

QY 11 LLAIVI--QFLGIFTNGIIVVNGIDILKHKMAPDLISCLAVSRIFLQLEFYVANI 68  
 DB 9 LLCVTVSEAVLGVDFTALANCMETAKKRLSKIGLITGLAISRIQVMIILQGM 68  
 QY 69 VIFIEFIMCSANCA-----ILFIELEMLATWLGVCYAKVASVHPFLTKM 123  
 DB 69 QVFF-PHILFEGNITYIYIWEFLNHSVWFATNLNLIYFPIKIANFSVFLMKSRVR 127  
 QY 124 KLVPMMLIGSLYVNICVF--HSKYAGFMPVFLRKFSQ--NATIOKEDTLAIOFIS 178  
 DB 128 --VFIFLSCGLTSLWLCFQPSKMLNNSKMYWGNTSMLOQOKNVLINQSLTNGITF 185  
 QY 179 FVAEFSVPLILFLFAVLLIFSLGRHROMKNTVAGSRVGRGAPISALISLILYF 238

DB 186 FI-----IVSLITCELLIVLMRIRQMSDGSGLRDNTBAHVAKMRLISFAVLFI 238  
 QY 239 SHCMIKVFLSLKFIH--RRFIFLFIYIGIYPSGHSILILINPKRLKONAKKFLHSHK 296  
 DB 239 LH-FVGLSIVLCFPLQNNLLEFITGLATCLYPCGHSITLILGNQLOKASLAKLOHLT 297  
 QY 297 CCQ 299  
 DB 298 CCE 300

RESULT 15  
 Q9JKT5 PRELIMINARY; PRT; 309 AA.  
 AC Q9JKT5;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 OS CANDIDATE TASTE RECEPTOR T2R9.  
 OC Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Adler E., Hoon M.A., Mueller K.L., Chandrashekar J., Ryba N.J.P.,  
 RA Zuker C.S.;  
 RT "A novel family of mammalian taste receptors."  
 RL Cell 100:693-702(2000).  
 DR EMBL: AF227146; AAF43919.1;  
 KW Receptor.  
 SQ SEQUENCE 309 AA; 35149 MW; 0FCE43EB9D3F0090 CRC64;

Query Match 20.8%; Score 317; DB 11; Length 309;  
 Best Local Similarity 29.1%; Pred. No. 8,5e-15;  
 Matches 91; Conservative 61; Mismatches 121; Indels 40; Gaps 8;

QY 1 MESHLLIYFLAVIQFLGIFTNGIIVVNGIDILKHKMAPDLISCLAVSRIFLQ 60  
 DB 1 MLSAAGILLSTAYVAGLGVGNTFIALVNCMDAKKRLSKIGFLGLATSRIFLYW 60  
 QY 61 FIFVNVIVIF-----IEFTMCSANCAILFIELEMLATWLGVCYAKVASVR 111  
 DB 61 ILIDAVAKLFPFGKYLSKLTREIISC-----IMTVNMMVFWFATLSIFFLKIANFS 115  
 QY 112 HPLFLTKMRISK-----LVPMMMLIGSLYSMICVHRSKYAGFNV--PYFLKRF 160  
 DB 116 HYFLMLKRTDKVFAFLMCLISMAL--SFSPYVKWKSMPKKNHGRNTSGTHWEKREF 173  
 QY 161 SONATIOKEDTLAIOFSPVAEFSVPLILFLFAVLLIFSLGRHROMKNTVAGSRVGR 220  
 DB 174 TSNVYL-----INIGVISL-----LIMTLFACELLILISLKHRSRONQSVSGFRDINT 221  
 QY 221 GAPISALISLFLILYSHCM-IVFLSLKFIHRFIFLFIYIGIYPSGHSILITL 279  
 DB 222 EAHVAIKFLLSIFLILFIYIGVAVELICMEIPENKLIIFGLTTASVPCCHSVILIL 281  
 QY 280 GNPRLKONAKKFL 292  
 DB 282 TNSQLKQAFVKYL 294

Search completed: March 15, 2001, 13:19:12  
 Job time: 1308 sec



THIS PAGE BLANK US

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 12:58:00 ; Search time 71.14 seconds  
(Without alignments)

135.731 Million cell updates/sec

Title: US-09-510-332-1

Perfect score: 1521

Sequence: 1 MESHLLITFLAVIQFLIG.....GNPKLKNKKFLHSKCCQ 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131	8.6	333	1 NU2M_APLI	P34849 apis mellif
2	130.5	8.6	348	1 OPSD_RABIT	P45912 oryctolagus
3	129.5	8.5	348	1 OPSD_SHEEP	P02700 ovis aries
4	129.5	8.5	354	1 OPSD_RAJER	P79863 raja eriac
5	128.5	8.4	348	1 OPSD_CRIGR	P28681 cricetus
6	128.5	8.4	348	1 OPSD_RAT	P51889 rattus norv
7	128.5	8.3	354	1 OPSD_BUFBU	P56514 bufu bufu
8	125.5	8.3	348	1 OPSD_MESBI	P62798 mesopiodon
9	125	8.2	348	1 OPSD_TURTR	P02798 tursiops tr
10	123.5	8.1	312	1 OZCL_HUMAN	P08100 homo sapien
11	123.5	8.1	348	1 OPSD_HUMAN	P02791 delphinus d
12	123	8.1	348	1 OPSD_DEIDE	P01493 homo sapien
13	122.5	8.1	346	1 GP41_HUMAN	P02794 phoca vitul
14	122.5	8.1	348	1 OPSD_HUMOV	P01866 sus scrofa
15	122.5	8.0	348	1 OPSD_PIG	P29867 drosophila
16	121.5	8.0	274	1 NU2M_PIG	P29866 macaca fasc
17	121.5	8.0	348	1 OPSD_MACFA	P15409 mus musculu
18	121.5	8.0	348	1 OPSD_MOUSE	P02796 trichechus
19	121.5	8.0	348	1 OPSD_TRIMA	P02792 globicephal
20	121	8.0	348	1 OPSD_GLOME	P51471 anolis caro
21	121	8.0	355	1 OPSB_ANOCA	P51750 borrelia bu
22	121	8.0	512	1 MVIN_BORBU	P09816 kaposi's sa
23	120.5	7.9	342	1 VG74_KSHV	P15529 homo sapien
24	120.5	7.9	346	1 GP42_HUMAN	P32308 canis famli
25	120.5	7.9	348	1 OPSD_CANFA	P09214 anguilla an
26	120.5	7.9	352	1 OPSD_ANGAN	P03459 scyllorhinu
27	120.5	7.9	354	1 OPSG_CARCA	P32311 carassius a
28	120	7.9	349	1 OPSG_CARAU	P22671 lampetra ja
29	119.5	7.9	353	1 GTR2_LAMJA	P11168 homo sapien
30	118.5	7.8	324	1 GTR2_HUMAN	P49219 xenopus lae
31	118	7.8	420	1 MLIC_XENLA	P02795 phoca groen
32	117.5	7.7	348	1 OPSD_PHOGR	P12336 rattus norv
33	117.5	7.7	522	1 GTR2_RAT	

34	117	7.7	354	1 OPSD_GALML	O93441 galeus mela
35	116.5	7.7	341	1 NU2M_DROME	P03896 drosophila
36	116.5	7.7	354	1 OPSD_ZEUPA	O42604 zeus faber
37	116.5	7.7	359	1 AC2R_BOVIN	P25104 bos taurus
38	116	7.6	353	1 OPSD_DIPVU	O99404 diploodus vu
39	115.5	7.6	348	1 OPSD_BOVIN	P02699 bos taurus
40	115.5	7.6	352	1 OPSD_ANOCA	P41591 anolis caro
41	115.5	7.6	355	1 OPSG_CHICK	P28683 gallus gall
42	114.5	7.5	353	1 OPSD_PETMA	O98980 petromyzon
43	114.5	7.5	354	1 OPSD_XENLA	P23403 xenopus lae
44	114.5	7.5	362	1 RODCL_RAT	O89039 rattus norv
45	114	7.5	289	1 OPSD_COTIN	O42330 cottocomeph

## ALIGNMENTS

RESULT 1  
ID NU2M\_APLI STANDARD: PRT: 333 AA.  
AC P34849; P92496; P92497; P92502; P92886; P92907; Q33785; Q33786;  
AC Q33787; Q33788; Q36885;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE NMDH-UBIQUINONE-OXIDOREDUCTASE CHAIN 2 (BC 1.6.5.3).  
GN N2.  
OS Apis mellifera ligustica (Common honeybee).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
OC Apoidea; Apidae; Apis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=THORAX;  
RX MEDLINE=93114603; PubMed=8417993;  
RA Crozier R.H., Crozier Y.C.;  
RT "The mitochondrial genome of the honeybee Apis mellifera: complete  
sequence and genome organization.";  
RT Genetics 133:97-117(1993).  
RN [2]  
RP SEQUENCE OF 1-211 FROM N.A.  
RC STRAIN=VARIOUS;  
RX MEDLINE=96364025; PubMed=8744768;  
RA Arias M.C., Sheppard W.S.;  
RT "Molecular phylogenetics of honey bee subspecies (Apis mellifera L.)  
inferred from mitochondrial DNA sequence.";  
RT Mol. Phylogenet. Evol. 5:557-566(1996).  
RN [3]  
RP SEQUENCE OF 60-307 FROM N.A.  
RC STRAIN=HAPLOTYPE 1 TO 11;  
RA Koulianos S., Crozier R.H.;  
RT Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.  
CC -!- CARAYINACTIVITY: NADH + UBIQUINONE - NAD(+) + UBIQUINOL.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC -!- INNER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: I06178; AAB96798.1; -  
CC EMBL: U35743; AAB38223.1; -  
CC EMBL: U35744; AAB38224.1; -  
CC EMBL: U35745; AAB38225.1; -  
CC EMBL: U35746; AAB38226.1; -  
CC EMBL: U35747; AAB38227.1; -  
CC EMBL: U35748; AAD12744.1; -





```

FT TRANSMEM 285 309 7 (POTENTIAL).
FT DOMAIN 310 348 CYTOPLASMIC.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT BINDING 296 296 RETINAL CHROMOPHORE.
FT BINDING 322 322 PALMITATE (BY SIMILARITY).
FT LIPID 323 323 PALMITATE (BY SIMILARITY).
FT LIPID 323 323 PALMITATE (BY SIMILARITY).
FT DISULFID 110 187 BY SIMILARITY.
FT MOD_RES 343 343 PHOSPHORYLATION (BY RK) (BY SIMILARITY).
SQ SEQUENCE 348 AA; 38994 MW; 09405AB5EA2E1E4 CRC64;

Query Match 8.6%; Score 130.5; DB 1; Length 348;
Best Local Similarity 21.8%; Pred. No. 0.02;
Matches 70; Conservative 64; Mismatches 128; Indels 59; Gaps 16;

OY 1 MESHLLIYFLAVIOFLGFTNGIIVVNGIDLIKRRKM-APDLLLSCLAVSRIFLQ 59
DB 39 MLAAYM---FLIVLGFPINFLVTV-----QHKRLTPLVNLTILMLVADLEFMV 87
OY 60 LFIPIYVNIIVIFIEIFMCSANCAILLFI---NELELM---LATMVGIFYCAVASVR 111
DB 88 LGFTTTLTSLHGYFVPGTCNVEGFATLGGELALMSLVLAIERVYVCKPMSNFR 147
OY 112 ---HPL---FTMLKMRISKLVPMMLIGSLLYV---SMICVFHSKYAGFMVPELRFES 161
DB 148 FENNAIMKVAFTWI-MALACAP-PIVGMSTRYIPEGMCCSGIDY-----YTLKPEVN 199
OY 162 QNATIQEDTLAIQIFSEVAESVLLIFLAVLLIFSLGHTQMNNTVAGSRVPRG 221
DB 200 NSEFV-----LYMFVHFHTIPLIIFFCYGLVFEVKEAAQODES-ATYKAEKE 249
OY 222 APISALLSTSLFLVESHCMKVFSLKFIIRRIPIF---FFLVIGIYSGHSLT 277
DB 250 VTRMVIIVYATLFCWVPASVAFIYFTHQSGNFGIIFMTIAFAFKSSSIY---NPVIY 306
OY 278 ILGNPKLONAKKFLHSKCC 298
DB 307 IMNKQF---RNCMLTITCC 323

RESULT 3
OPSD_SHEEP STANDARD: PRT; 348 AA.
ID P02700;
AC 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
RN [1].
RP SEQUENCE.
RA Pappin D.J.C., Elipoulos E., Brett M., Findlay J.B.C.;
RT "A structural model for ovine rhodopsin."
RL Int. J. Biol. Macromol. 6:73-76(1984).
RN [2].
RP SEQUENCE OF 1-111 AND 144-239.
RA MEDLINE=83282605; PubMed=6224479;
RX Brett M., Findlay J.B.C.;
RT "Isolation and characterization of the CNBR peptides from the
RT proteolytically derived N-terminal fragment of ovine opsin."
RL Biochem. J. 211:661-670(1983).
RN [3].
RP SEQUENCE OF 240-348.
RX MEDLINE=82013638; PubMed=7278988;
RA Findlay J.B.C., Brett M., Pappin D.J.C.;
RT "Primary structure of C-terminal functional sites in ovine
RT rhodopsin."

```

```

RL Nature 293:314-316(1981).
RN [4].
RP RETINAL BINDING SITE.
RX MEDLINE=84178280; PubMed=6370231;
RA Pappin D.J.C., Findlay J.B.C.;
RT "Sequence variability in the retinal-attachment domain of mammalian
RT rhodopsins."
RL Biochem. J. 217:605-613(1984).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED.
CC -1- MISCELLANEOUS: THIS RHODOPSIN HAS AN ABSORPTION MAXIMA AT 495 NM.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
DR PTR: A30407; OOSH.
DR GCRDB: GCR_0194.
DR INTERPRO: IPR000276;
DR INTERPRO: IPR000732;
DR INTERPRO: IPR001760;
DR PFAM: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCRHHODOPSN.
DR PRINTS: PR00238; OPSIN.
DR PRINTS: PR00579; RHODOPSIN.
DR PROSITE: PS00237; G-PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS50262; G-PROTEIN_RECEP_F1_2; 1.
DR PROSITE: PS00238; OPSIN; 1.
KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
FT DOMAIN 1 36 1 (POTENTIAL).
FT TRANSMEM 37 61 1 (POTENTIAL).
FT DOMAIN 62 73 2 (POTENTIAL).
FT TRANSMEM 74 98 3 (POTENTIAL).
FT DOMAIN 99 113 3 (POTENTIAL).
FT TRANSMEM 114 133 4 (POTENTIAL).
FT DOMAIN 134 152 4 (POTENTIAL).
FT TRANSMEM 153 176 5 (POTENTIAL).
FT DOMAIN 177 202 5 (POTENTIAL).
FT TRANSMEM 203 230 6 (POTENTIAL).
FT DOMAIN 231 252 6 (POTENTIAL).
FT TRANSMEM 253 276 7 (POTENTIAL).
FT DOMAIN 277 284 7 (POTENTIAL).
FT TRANSMEM 285 309 7 (POTENTIAL).
FT DOMAIN 310 348 7 (POTENTIAL).
FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT BINDING 296 296 RETINAL CHROMOPHORE.
FT LIPID 322 322 PALMITATE (BY SIMILARITY).
FT LIPID 323 323 PALMITATE (BY SIMILARITY).
FT DISULFID 110 187 BY SIMILARITY.
FT MOD_RES 343 343 PHOSPHORYLATION (BY RK) (BY SIMILARITY).
SQ SEQUENCE 348 AA; 38991 MW; AAFD6F0DA8BAE5 CRC64;

Query Match 8.5%; Score 129.5; DB 1; Length 348;
Best Local Similarity 21.1%; Pred. No. 0.023;
Matches 69; Conservative 62; Mismatches 125; Indels 71; Gaps 16;

OY 1 MESHLLIYFLAVIOFLGFTNGIIVVNGIDLIKRRKM-APDLLLSCLAVSRIFLQ 59
DB 39 MLAAYM---FLIVLGFPINFLVTV-----QHKRLTPLVNLTILMLVADLEFMV 87
OY 60 LFIPIYVNIIVIFIEIFMCSANCAILLFI---NELELM---LATMVGIFYCAVASVR 111
DB 88 LGFTTTLTSLHGYFVPGTCNVEGFATLGGELALMSLVLAIERVYVCKPMSNFR 147
OY 112 ---HPL---FTMLKMRISKLVPMMLIGSLLYV---SMICVFHSKYAGFMVPELRF 157
DB 148 FENNAIMKVAFTWIMALACAPPIVGM-----SRY-----IPQMGCCSC 187

```

QY 158 --KFSQONATICKEDTLAIOIFSVFAESVPLILFLFAVLLIFSLGHTROMNTVAGS 215  
 DB 188 GALTFLYLRKEINNE---SVIYIMFVHRSIPILVIFPCYGGOLVFYKKAQAQOQES-RTT 243  
 QY 216 RYFGKCAPLSALLSLIFLIFSHCMIKVLSLKFHIREPL-FFLLVIGIYPS 271  
 DB 244 QAKAEVTRMVIIMVIAFLICMLPYAGVAFYFTFHQSDFCPIFTIPAFKSSSV-- 301  
 QY 272 GHSLLILGNPKLONAKFELHSCC 298  
 DB 302 -NPVIYIMNKKF-----RNCMLTLTLC 323

RESULT 4  
 OPSD\_RAJER STANDARD; PRT: 354 AA.  
 AC P79863;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE RHODOPSIN.  
 GN RHO.  
 OS Raja erinacea (Little skate).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 OC Elasmobranchii; Squalia; Hypnosqualia; Pristiogadidae; Batoidae;  
 OC Rajiformes; Rajidae; Raja.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-RETINA;  
 RX MEDLINE-97398145; PubMed-9256070;  
 RA O'Brien J., Rippes H., Al-Ubaidi M.R.;  
 RL "Molecular cloning of a rod opsin cDNA from the skate retina.";  
 CC Gene 193:141-150(1997).  
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT  
 CC MEDATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY  
 CC LINKED TO CIS-RETINAL.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES  
 CC VISION IN DIM LIGHT.  
 CC -1- PFM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY  
 CC BE PHOSPHORYLATED.  
 CC -1- MISCELLANEOUS: THIS RHODOPSIN HAS AN ABSORPTION MAXIMA AT 501 NM.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC OPSIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U01514; AAC60251.1; --  
 CC GCRDB: GCR\_1485; --  
 CC DR INTERPRO: IPR000276; --  
 CC DR INTERPRO: IPR000732; --  
 CC DR INTERPRO: IPR001760; --  
 CC PFAM: PF00001; 7tm.1; 1.  
 CC DR PRINTS: PR00237; GPCRHOOPS.  
 CC DR PRINTS: PR00238; OPSIN.  
 CC DR PRINTS: PR00579; RHODOPSIN.  
 CC DR PROSITE: PS00237; G-PROTEIN-RECEP\_F1\_1; 1.  
 CC DR PROSITE: PS00262; G-PROTEIN-RECEP\_F1\_2; 1.  
 CC DR PROSITE: PS00238; OPSIN; 1.  
 CC KW Photoreceptor. Retinal protein. Transmembrane; Glycoprotein; Vision;  
 CC Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.  
 FT DOMAIN 1 36  
 FT TRANSMEM 37 61  
 FT DOMAIN 62 73  
 FT TRANSMEM 74 98  
 FT DOMAIN 99 113  
 CC EXTRACELLULAR.  
 CC 1 (POTENTIAL).  
 CC 2 (POTENTIAL).  
 CC CYTOPLASMIC.  
 CC 2 (POTENTIAL).  
 CC EXTRACELLULAR.

FT TRANSMEM 114 133  
 FT DOMAIN 134 152  
 FT TRANSMEM 153 176  
 FT DOMAIN 177 202  
 FT TRANSMEM 203 230  
 FT DOMAIN 231 252  
 FT TRANSMEM 253 276  
 FT DOMAIN 277 284  
 FT TRANSMEM 285 309  
 FT DOMAIN 310 354  
 FT CARBOHYD 2 2  
 FT CARBOHYD 15 15  
 FT BINDING 296 296  
 FT LIPID 322 322  
 FT DISULFID 110 187  
 SQ SEQUENCE 354 AA; 39697 MW; 1ACBB918ADA214C2 CRC64;

Query Match 8.5%; Score 129.5; DB 1; Length 354;  
 Best Local Similarity 22.2%; Pred. No. 0.023; Indels 83; Gaps 18;  
 Matches 70; Conservative 56; Mismatches 107;

QY 8 IYFLAVIOELGIFNGIIVVNGIDLIKRRK-APLDLLSCLAVSRIFQLFFIYVN 66  
 DB 44 MFPLI-----LTGLPVNPLFLVLT-----IQHKKLRQPLNVIILNLAVSDLFVFGFTTT 94  
 QY 67 VIYIFIEPIFMCANCAILFT-----NELEMLATWLVG---FYCAKVASVR---HPL 114  
 DB 95 IITSMNGYFIEFGAGNFEQFEFATLGEVGLMCLVLAIERVYVCKPMANFRFGSOHAI 154  
 QY 115 ----FTLMKRRIS----KIVPMIIGSLIYVSMICVFHSKYVGFVYPLR-----KFS 161  
 DB 155 IGVEYFMI-MALSCAPPLVGM-----SRY-----IPESLQSCGVDYIT 193  
 QY 162 QANTTKEDTLAIOIFSVFAESVPLILFLFAVLLIFSLGHTROMNTVAGSRVPGRG 221  
 DB 194 MKREVNE---SVIYIMFVHRSIPILVIFPCYGGOLVFYKKAQAQOQESSTORAE-RE 249  
 QY 222 APISALLSLIFLIFSHCMIKVLSLKFHIREPL-FFLLVIGIYPSG- 272  
 DB 250 VPRMVIIMVIAFLICMLPYAGVAFYFTFHQSDFCPIFTIPAFKSSSV 298  
 QY 273 ---HSLLILGNPKL 285  
 DB 299 AVYNPLIYIMNKKFR 314

RESULT 5  
 OPSD\_CRIGR STANDARD; PRT: 348 AA.  
 AC P28681;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE RHODOPSIN.  
 GN RHO.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetulus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-OVARY;  
 RX MEDLINE-92219256; PubMed-1560457;  
 RA Gale J.M., Tobey R.A., D'Anna A.;  
 RL "Localization and DNA sequence of a replication origin in the  
 RT rhodopsin gene locus of Chinese hamster cells.";  
 J. Mol. Biol. 224:343-358(1992).  
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT  
 CC MEDATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY  
 CC LINKED TO CIS-RETINAL.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES

```

Oy 162 ONATOKEEDFLAIQIFSPVAESVPLLIPEFAVLIIIFSLGHTTQMRTVYGSVPWG 221
Db 200 NESFV-----IYMFVFHTPIPLIVIFECYGGVLFTVKEAQAQDS ATTOKAEK 249
Oy 222 APISALLSIIFLIYVSHCMIKVFSSLKFEHRRRIFL----FFLIIVGIVPSGSLIL 277
Db 250 VTRNVILNVVFFLIICWPPYAGVAFYFTTHGSGNFCPIFMTLPFAFFAKSSSIY---NPVIY 306
Oy 278 ILGNPKLKNAKKFKLHKKCC 298
Db 307 IMANKQF-----RNCMLTTLCC 323

RESULT 6
OPSD_RAT      OPSD_RAT      STANDARD;      PRT;      348 AA.
AC      P51489;
Dr 01-OCT-1996 (Rel. 34, Created)
Dr 01-OCT-1996 (Rel. 34, Last sequence update)
Dr 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN
RN SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-RETINAL ROD CELL;
RX MEDLINE-95383121; PubMed-7654522;
RT Barstable C.J., Morabito M.A.;
RT *Isolation and coding sequence of the rat rod opsin gene.*;
RL J. Mol. Neurosci. 5:207-209(1994).
RN
RN SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-RETINA;
RA Huber A., Baker B.B., Sander P., Gerdon G., Paulsen R.,
RA Williams T.P.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED.
CC -1- MISCELLANEOUS: THIS RHODOPSIN HAS AN ABSORPTION MAXIMA AT 495 NM.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC
CC -----
CC THIS SMIS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
Dr EMBL; U22180; AAA84439.1; -
Dr EMBL; Z46957; CAA87081.1; -
Dr GCRDB; GCR.1458; -
Dr GCRDB; GCR.1458; -
Dr INTERPRO; IPR000276; -
Dr INTERPRO; IPR000732; -
Dr INTERPRO; IPR001760; -
Dr PFAM; PF00001; 7tm_1; 1.
Dr PRINTS; PR00237; GPCRHOPOPSN.
Dr PRINTS; PR00238; OPSIN.
Dr PRINTS; PR00579; RHODOPSIN.
Dr PROSITE; PS00237; G-PROTEIN_RECEP_FL_1; 1.
Dr PROSITE; PS50262; G-PROTEIN_RECEP_FL_2; 1.
Dr PROSITE; PS00238; OPSIN; 1.
Dr Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW

```



```

OY 113 PL-----FTWLMKRIKSLVPMNIIISGLIYV--SMICVFHSHKAGENVYFLRKFFSONATI 166
DB 153 AVMGVAFTWI-MALSCAVP-PLIGMSRYIEGMOCSGVDY-----YTLKPEVNNESFY 204
OY 167 QKEPTLAIQISFVAERSPELLIFLFAVLLIFSLGRHTROMNTVAGSVPGCAPISA 226
DB 205 -----IIMFVVFHTPIITIIIFFCYGRVCTVKNAAAOQDS-ATQKAKEVTRMV 254
OY 227 LLSLISFLILYFSHQMKVLSLKFHIRFEL----FPLIVGITYPSGSHLILIGNP 282
DB 255 IIMVFEILCWVYASVAFEFISNGSEFGPIFMTVPAFEAKSSSIY---NPVIYIM--- 308
OY 283 KIKONAKKFLHSHKCC 298
DB 309 -LNKQFRNCMTITLLCC 323

RESULT 8
OPSD_MESBI STANDARD: PRT: 348 AA.
ID OPSD_MESBI 062793:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO.
OS Mesopodon bidens.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Ziphiidae;
OC Mesopodon.
RN [1]
RP SEQUENCE FROM N.A.
RA Pasick J.I., Robinson P.R.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF055316; AAC12763.1; -
CC DR INTERPRO: IPR000276; -
CC DR INTERPRO: IPR000732; -
CC DR INTERPRO: IPR001760; -
CC PFAM: PF00001; 7tm_1; 1.
CC DR PRINTS: PR00237; GPCRHHODOPSIN.
CC DR PRINTS: PR00238; OPSIN.
CC DR PRINTS: PR00579; RHODOPSIN.
CC DR PROSITE: PS00237; G-PROTEIN_RECEP_F1_1; 1.
CC DR PROSITE: PS00262; G-PROTEIN_RECEP_F1_2; 1.
CC DR PROSITE: PS00238; OPSIN; 1.
CC KW Photorceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
CC Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor;
CC Acetylation.
CC KM DOMAIN 1 36 EXTRACELLULAR.
CC FT DOMAIN 1 61 1 (POTENTIAL).
CC FT TRANSMEM 62 73 CYTOPLASMIC.
CC FT TRANSMEM 74 98 2 (POTENTIAL).
CC FT DOMAIN 99 113 EXTRACELLULAR.

```

```

FT TRANSMEM 114 133 3 (POTENTIAL).
FT DOMAIN 134 152 CYTOPLASMIC.
FT TRANSMEM 153 176 4 (POTENTIAL).
FT DOMAIN 177 202 EXTRACELLULAR.
FT TRANSMEM 203 230 5 (POTENTIAL).
FT DOMAIN 231 252 CYTOPLASMIC.
FT TRANSMEM 253 276 6 (POTENTIAL).
FT DOMAIN 277 284 EXTRACELLULAR.
FT TRANSMEM 285 309 7 (POTENTIAL).
FT DOMAIN 310 348 CYTOPLASMIC.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 110 187 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT BINDING 296 296 BY SIMILARITY.
FT LIPID 322 322 RETINAL CHROMOPHORE.
FT LIPID 323 323 PALMITATE (BY SIMILARITY).
FT MOD_RES 343 343 PHOSPHORYLATION (BY RK) (BY SIMILARITY).
SQ SEQUENCE 348 AA; 39023 MW; F83010DC40C450D5 CRC64;

```

Query Match 8.3%; Score 125.5; DB 1; Length 348;  
 Best Local Similarity 20.8%; Pred. No. 0.042; Indels 67; Gaps 16;  
 Matches 67; Conservative 63; Mismatches 125;

```

OY 9 YELLAAYOPL--IGITNGIIVVNGIDILKRRKM-APDLLSLCLAVSRIFLQPIFY 64
DB 37 FSVLAAYMFLIMLGPPINFLTYT---VQHKRLTPILNLYLMLAVANLEFVIGFT 92
OY 65 VNVIVIFIEFIMCSANCAILLFT---NELELM---LATWIGVFYCAVASVR---H 112
DB 93 TTYLTSMAHYTIFGPTGCLNEGFEATLGEIALLSLVLAIERVYVVCXKMSNFRGENH 152
OY 113 PL-----FTW---LKMRIKSLVPMNIIISGLIYVSMICVFHSHKAGENVYFLR-----KFP 160
DB 153 AIMGLATLWIMALLCAAPPLVGM-----SRV-----IPEGMQCSGVDY 192
OY 161 SONATIQEPTLAIQISFVAERSPELLIFLFAVLLIFSLGRHTROMNTVAGSVRPGR 220
DB 193 TSPPEVNN-----SFVYVVFVHFSPVPIIFFCYGOALVFYVKNAAAOQDS-ATQKAER 248
OY 221 GAPISALLISFLILYFSHQMKVLSLKFHIRFEL----FPLIVGITYPSGSHL 276
DB 249 EYTRVYVIMVAVFLICWVYASVAFEFITHOGSNFGPIFMTIPSFPAKSSSIY---NPVI 305
OY 277 IILGNPKLQNAKKFLHSHKCC 298
DB 306 YIMMKNQF-----RNCMLTITLLCC 323

```

```

RESULT 9
OPSD_TURTR STANDARD: PRT: 348 AA.
ID OPSD_TURTR 062798:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO.
OS Tursiops truncatus (Atlantic bottle-nosed dolphin).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;
OC Tursiops.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98129138; PubMed-9471225;
RA Pasick J.I., Robinson P.R.;
RL Biochemistry 37:433-438(1998).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

```

CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES  
 CC VISION IN DIM LIGHT.  
 CC -1- PTM: SOME OR ALL OF THE CARBOXY-TERMINAL SER OR THR RESIDUES MAY  
 CC BE PHOSPHORYLATED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC OPSIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF055456; AAC12940.1; -  
 CC INTERPRO: IPR000276; -  
 CC INTERPRO: IPR000732; -  
 CC INTERPRO: IPR001760; -  
 CC PFAM: PF00001; 7tm1.1;  
 CC PRINTS: PRO0237; GPCRHDOPSN.  
 CC PRINTS: PRO0238; OPSIN.  
 CC PROSITE: PS00237; G-PROTEIN\_RECEP\_F1\_1;  
 CC PROSITE: PS50262; G-PROTEIN\_RECEP\_F1\_2; 1.  
 CC PROSITE: PS00238; OPSIN; 1.  
 CC DR Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;  
 KW phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor;  
 KW Acetylation.  
 FT DOMAIN 1 36 EXTRACELLULAR.  
 FT TRANSSEM 37 61 1 (POTENTIAL).  
 FT DOMAIN 62 73 CYTOPLASMIC.  
 FT TRANSSEM 74 98 2 (POTENTIAL).  
 FT DOMAIN 99 113 EXTRACELLULAR.  
 FT TRANSSEM 114 133 3 (POTENTIAL).  
 FT DOMAIN 134 152 CYTOPLASMIC.  
 FT TRANSSEM 153 176 4 (POTENTIAL).  
 FT DOMAIN 177 202 5 (POTENTIAL).  
 FT TRANSSEM 203 230 6 (POTENTIAL).  
 FT DOMAIN 231 252 7 (POTENTIAL).  
 FT TRANSSEM 253 276 8 (POTENTIAL).  
 FT DOMAIN 277 284 EXTRACELLULAR.  
 FT TRANSSEM 285 309 7 (POTENTIAL).  
 FT DOMAIN 310 348 7 (POTENTIAL).  
 FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 FT CARBOHYD 15 2 N-LINKED (GLCNAC. . .) (BY SIMILARITY).  
 FT DISULFID 110 187 N-LINKED (GLCNAC. . .) (BY SIMILARITY).  
 FT BINDING 296 296 RETINAL CHROMOPHORE.  
 FT LIPID 322 322 PALMITATE (BY SIMILARITY).  
 FT LIPID 322 322 PALMITATE (BY SIMILARITY).  
 FT MOD\_RES 343 343 PHOSPHORYLATION (BY RK) (BY SIMILARITY).  
 SQ SEQUENCE 348 AA; 39093 MW; 32605DE5ED94723A CRC64;

Query Match 8.2%; Score 125; DB 1; Length 348;  
 Best Local Similarity 20.5%; Pred. No. 0.046;  
 Matches 64; Conservative 61; Mismatches 131; Indels 56; Gaps 14;

QY 10 FLAVIOLFLGFTNGIIVVNGIDILKHKM-APDLLSLCLAVSRITLQFLIPVANI 68  
 DB 45 FLIVLGFPTNFTLVVY-----QHKKLTPLNTYLLNLAANLPMVGGFTTLY 96  
 QY 69 VIFIEFIMCSANCAILLFT---NELEIM---LATWLVGYCAKVASR---HPL-- 114  
 DB 97 TSLHAFVEFGPTCNLEGFATLGEIALSLVLAIEYVVVCKPMNSFRGENHAING 156  
 QY 115 --ETWIKMRISKVPMIIGSLLY--SMICVPHSKYAGFMPYLKRFESONATQED 170  
 DB 157 LALTWTMAACAAP--LVGWSRYLPEGMOC-----SCGIDYTSR-----QEVN 199  
 QY 171 TLAQIFSEVAEFSVPLFLFAVLLIFSLGRHTROMNTAVAGSRVGRGAPISALLSI 230  
 . . . 6 . . . . .

DB 200 NESFVIMFVHFTPLVLIIFCYGOLVETKBAADQOES-ATOKAREYTRVIMV 258  
 QY 231 LSFLLFYSCHMIKVFSLSKPIHRPIFL---FPLVIGYPSGSHLLILGNKLMQ 286  
 DB 259 VAFLLCWVPYASVAFYFTHGSDPGIFMTIFSPFAKSSII---NPVIYIMNKQF-- 313  
 QY 287 NAKKFLHNSCC 298  
 DB 314 --RNCMLTTLCC 323  
 RESULT 10  
 ID 02CL\_HUMAN STANDARD; PRT; 312 AA.  
 AC 095371;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE OLFATORY RECEPTOR 2CL.  
 GN OR2C1.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Centola M., Chen X., Sood R., Deng Z., Aksentjevich I., Blake T.,  
 RA Ricke D., Chen X., Wood G., Zaks N., Richards N., Krizman D.,  
 RA Mansfield E., Apostolou S., Liu J., Shafian N., Vedula A., Hamon M.,  
 RA Cercer A., Kahan T., Guncuo D., Callen D.F., Richards R.I.,  
 RA Moysis R.K., Doggett N.A., Collins F.S., Liu P.P.,  
 RA Fischel-Ghodsian N., Kastner D.L.;  
 RT \*Construction of an approximately 700 kb transcript map around the  
 RL familial mediterranean fever locus on human chromosome 16p13.3.\*;  
 CC Genome Res. 9:1-21(1998)  
 CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF098664; AAC83557.1; -  
 CC INTERPRO: IPR000276; -  
 CC INTERPRO: IPR000725; -  
 CC PFAM: PF00001; 7tm1.1;  
 CC PRINTS: PRO0237; GPCRHDOPSN.  
 CC PRINTS: PRO0245; OLFATORYR.  
 CC PROSITE: PS00237; G-PROTEIN\_RECEP\_F1\_1;  
 CC PROSITE: PS50262; G-PROTEIN\_RECEP\_F1\_2; 1.  
 CC DR G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW Multigene family; Olfaction.  
 FT DOMAIN 1 25 EXTRACELLULAR (POTENTIAL).  
 FT TRANSSEM 26 49 1 (POTENTIAL).  
 FT DOMAIN 50 57 CYTOPLASMIC (POTENTIAL).  
 FT TRANSSEM 58 79 2 (POTENTIAL).  
 FT DOMAIN 80 100 EXTRACELLULAR (POTENTIAL).  
 FT TRANSSEM 101 120 3 (POTENTIAL).  
 FT DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).  
 FT TRANSSEM 140 158 4 (POTENTIAL).  
 FT DOMAIN 159 196 EXTRACELLULAR (POTENTIAL).  
 FT TRANSSEM 197 219 5 (POTENTIAL).  
 FT DOMAIN 220 236 CYTOPLASMIC (POTENTIAL).  
 FT TRANSSEM 237 259 6 (POTENTIAL).  
 FT DOMAIN 260 272 EXTRACELLULAR (POTENTIAL).  
 FT TRANSSEM 273 292 7 (POTENTIAL).  
 FT DOMAIN 293 312 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 97 189 BY SIMILARITY.  
 FT CARBOHYD 5 N-LINKED (GLCNAC. . .) (POTENTIAL).



RP VARIANTS ADRP A-51; I-104; R-106; G-135; S-140; E-188; M-209 & R-211.  
 RX MEDLINE-93304432; PubMed-8317502;  
 RA Macke J.P., Davenport C.M., Jacobson S.G., Hennessey J.C.,  
 RA Gonzalez-Fernandez F., Conway B.P., Heckenlively J., Palmer R.,  
 RA Maumenee I.H., Steving P., Gouras P., Good W., Nathans J.;  
 RT "Identification of novel rhodopsin mutations responsible for  
 RT retinitis pigmentosa: implications for the structure and function of  
 RT rhodopsin."  
 RL Am. J. Hum. Genet. 53:80-89(1993).  
 RN [15]  
 RP VARIANT ADRP SER-15.  
 RX MEDLINE-93357759; PubMed-8353500;  
 RA Kranich H., Bartkowiak S., Denton M.J., Krey S., Dickinson P.,  
 RA Duvigneau C., Gal A.;  
 RT "Autosomal dominant 'sector' retinitis pigmentosa due to a point  
 RT mutation predicting an Asp-15-Ser substitution of rhodopsin."  
 RL Hum. Mol. Genet. 2:813-814(1993).  
 RN [16]  
 RP VARIANT CSNB4 GLU-292.  
 RX MEDLINE-93364423; PubMed-8358437;  
 RA Dryja T.P., Berson E.L., Rao V.R., Orlan D.D.;  
 RT "Heterozygous missense mutation in the rhodopsin gene as a cause of  
 RT congenital stationary night blindness."  
 RL Nat. Genet. 4:280-283(1993).  
 RN [17]  
 RP VARIANTS ADRP  
 RX MEDLINE-94375083; PubMed-8088850;  
 RA Valtinathan R., Berson E.L., Dryja T.P.;  
 RT "Further screening of the rhodopsin gene in patients with autosomal  
 RT dominant retinitis pigmentosa."  
 RL Genomics 21:461-463(1994).  
 RN [18]  
 RP VARIANT ADRP THR-44.  
 RX MEDLINE-94357587; PubMed-8076945;  
 RA Reig C., Antich J., Gean E., Garcia-Sandoval B., Ramos C., Ayuso C.,  
 RA Cardello M.;  
 RT "Identification of a novel rhodopsin mutation (Met-44-Thr) in a  
 RT simplex case of retinitis pigmentosa."  
 RL Hum. Genet. 94:283-286(1994).  
 RN [19]  
 RP VARIANTS ADRP PHE-110; PRO-131 AND VAL-164.  
 RX MEDLINE-95072600; PubMed-7981701;  
 RA Fuchs S., Kranich H., Denton M.J., Zrenner E., Bhattacharya S.S.,  
 RA Humphries P., Gal A.;  
 RT "Three novel rhodopsin mutations (C110P, L131P, A164V) in patients  
 RT with autosomal dominant retinitis pigmentosa."  
 RL Hum. Mol. Genet. 3:1203-1203(1994).  
 RN [20]  
 RP VARIANT ADRP GLN-171.  
 RX MEDLINE-9507852; PubMed-7987326;  
 RA Antinolo G., Sanchez B., Borrego S., Rueda T., Chaparro P.,  
 RA Cabeza J.C.;  
 RT "Identification of a new mutation at codon 171 of rhodopsin gene  
 RT causing autosomal dominant retinitis pigmentosa."  
 RL Hum. Mol. Genet. 3:1421-1421(1994).  
 RN [21]  
 RP VARIANTS ADRP PHE-127; PRO-131; ASN-178; ARG-267 AND ARG-297.  
 RX MEDLINE-9507858; PubMed-7987331;  
 RA Souied E., Gerber S., Rozet J.M., Bonneau D., Duffier J.L., Ghazi I.,  
 RA Philip N., Soubriere G., Coscas G., Munnich A.;  
 RT "Five novel missense mutations of the rhodopsin gene in autosomal  
 RT dominant retinitis pigmentosa."  
 RL Hum. Mol. Genet. 3:1433-1434(1994).  
 RN [22]  
 RP VARIANTS ADRP ARG-40 AND LYS-216.  
 RX MEDLINE-94362717; PubMed-8081400;  
 RA Al-Maghrabi M., Inglehearn C., Lunt P., Jay M., Bird A.,  
 RA Bhattacharya S.;  
 RT "Two new rhodopsin transversion mutations (U40R; M216K) in families  
 RT with autosomal dominant retinitis pigmentosa."  
 RL Hum. Mutat. 3:409-410(1994).  
 RN [23]  
 RP VARIANT ADRP LEU-345.

RX MEDLINE-94321123; PubMed-8045708;  
 RA Rosas D.J., Roman A.J., Weisbrod P., Macke J.P., Nathans J.;  
 RT "Autosomal dominant retinitis pigmentosa in a large family: a  
 RT clinical and molecular genetic study."  
 RL Invest. Ophthalmol. Vis. Sci. 35:3134-3144(1994).  
 RN [24]  
 RP VARIANT ADRP LYS-150.  
 RX MEDLINE-95078913; PubMed-7987385;  
 RA Kumaramanickavel G., Maw M., Denton M.J., John S., Srikumari C.R.,  
 RA Orth U., Oehlmann R., Gal A.;  
 RT "Missense rhodopsin mutation in a family with recessive RP."  
 RL Nat. Genet. 8:10-11(1994).  
 RN [25]  
 RP VARIANT ADRP ALA-347.  
 RX MEDLINE-95359993; PubMed-7633434;

Query Match 8.1%; Score 123.5; DB 1; Length 348;  
 Best Local Similarity 20.8%; Pred. No. 0.058;  
 Matches 68; Conservative 63; Mismatches 125; Indels 71; Gaps 16;

OY 1 MESHLLIYFLAVIQLGIFNGIIVVNGIDIKHRKM-APDLISCLAVSRIFLQ 59  
 DB 39 MAAVM---FLIVLGFPINFLVTV-----QHKKLRPLNYILNLAVADLEAV 87  
 OY 60 LFIFFVNVVIFEFIFIKSANCALILEFT---NELELM---LATWIGFYCARVASVR 111  
 DB 88 LGGFTSTLTSLHGYFVEGPTGCGNEGFATLGELTALMSLVLAIERVYVVCCKPMSNR 147  
 OY 112 ---HPL---FLW---LKMRSIKLVPMMILSLLYVSMICVFSKYGVMVPLR--- 157  
 DB 148 FGENHAIMGVAFWTWVALACAPPLAGW-----SRY-----IPELDQSC 187  
 OY 158 --KEFSQNTICKEDTLAQIFSPAERSVPLILFLVALLISLGRHTQMRNTVAGS 215  
 DB 188 GIDYTLAKEPVNNE---SVVIYFVHFTIPMIITFFCYGQLVFKVBAADQOES-ATY 243  
 OY 216 RVPGRCAPISALSLISFLIFESHCMIKVFLSKFPIHRIRFL---FLVIAGIYPS 271  
 DB 244 OKAEKEVTRMVIIMVIAFLICWVPYASVAFYFTHQSGNFGPIFTIAPFAKSAAYT-- 301  
 OY 272 GHSLLILGPKLKNOKAKFLIHSKCC 298  
 DB 302 -NPVIYIMNKP---RNCMTLTTC 323

RESULT 12  
 OPSD\_DELDE STANDARD; PRT; 348 AA.  
 AC 062791;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE RHODOPSIN.  
 GN RHO.  
 OS Delphinus delphis (Saddleback dolphin) (Black sea dolphin).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;  
 CC Delphinus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Fastick J.I., Robinson P.R.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT  
 CC LINKED TO CIS-RETINAL.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES  
 CC VISION IN DIM LIGHT.  
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY  
 CC BE PHOSPHORYLATED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC OPSIN SUBFAMILY.





```

Db 76 ANGMHMPLEFLCPLSG--FIFETIYL-TALFLAASIERFLSVANHL--WYKTRRLG 130
Oy 122 ---ISKLVPMWMLG---SLAVSMIC--VPHSKYAGFVWPFLEKFRSONATIO---KED 170
Db 131 OAGLVSAVCMLASHACGVAVVIEPSGDISHSQ-----GTNCTCYLERRKD 176
Oy 171 TLAT---QIFSPAESVPLLEFLAVLLIFSLGR---HTROMRVAGSRVPGCAP 223
Db 177 QLATLLPVLRLMAVAVLFWPLITSYCSRLWMLLGRSGRRRR-----222
Oy 224 ISALL--SLSEFLIX-----C-----MKVFLSLKFRHREFLF 261
Db 223 VAGLLATLLNPLVCFPGYVSHVGYIGESPPAMRIYVTLTLNLCVDPEVYF 278

RESULT 14
OPSD_PIG 14
ID OPSD_PIG 14 STANDARD: PRT: 348 AA.
AC 062754:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO.
OS Phoca vitulina (Harbor seal).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Plinipedia; Phocidae; Phoca.
RN [1]
RP SEQUENCE FROM N.A.
RA Fastig J.I., Robinson P.R.:
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OR G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF05317; AAC12764.1; -
CC DR INTERPRO: IPR000276; -
CC DR INTERPRO: IPR000732; -
CC DR INTERPRO: IPR001760; -
CC DR PFAM: PF00001; 7tm_1; 1.
CC DR PRINTS: PR00237; GPCRHHODOPSIN.
CC DR PRINTS: PR00238; OPSIN.
CC DR PRINTS: PR00579; RHODOPSIN.
CC DR PROSITE: PS00237; G-PROTEIN_RECIP_F1_1; 1.
CC DR PROSITE: PS00262; G-PROTEIN_RECIP_F1_2; 1.
CC DR PROSITE: PS00238; OPSIN; 1.
CC DR PROSITE: PS00238; OPSIN; 1.
CC DR Phocidae; Retinal protein; Transmembrane; Glycoprotein; Vision;
CC Acetylation.
CC KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor;
CC Acetylation.
CC FT DOMAIN 1 36 EXTRACELLULAR.
CC FT TRANSMEM 37 61 1 (POTENTIAL).
CC FT DOMAIN 62 73 CYTOPLASMIC.
CC FT TRANSMEM 74 98 2 (POTENTIAL).
CC FT DOMAIN 99 113 EXTRACELLULAR.
CC FT TRANSMEM 114 133 3 (POTENTIAL).
CC FT DOMAIN 134 152 CYTOPLASMIC.
CC FT TRANSMEM 153 176 4 (POTENTIAL).
CC FT DOMAIN 177 202 EXTRACELLULAR.

```

```

FT TRANSMEM 203 230 5 (POTENTIAL).
FT DOMAIN 231 252 CYTOPLASMIC.
FT TRANSMEM 253 276 6 (POTENTIAL).
FT DOMAIN 277 284 EXTRACELLULAR.
FT TRANSMEM 285 309 7 (POTENTIAL).
FT DOMAIN 310 348 CYTOPLASMIC.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 110 187 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT BINDING 296 296 RETINAL CHROMOPHORE.
FT LIPID 322 322 PALMITATE (BY SIMILARITY).
FT LIPID 323 323 PHOSPHORYLATION (BY RK) (BY SIMILARITY).
FT MOD_RES 343 343 PALMITATE (BY SIMILARITY).
SQ SEQUENCE 348 AA: 38973 MW: 8519CDBCEEFEB8 CRC64;

Query Match 8.1%; Score 122.5; DB 1; Length 348;
Best Local Similarity 19.9%; Pred. No. 0.067;
Matches 65; Conservative 66; Mismatches 125; Indels 71; Gaps 16;

Oy 1 MESHLLIYELLAVIOFLGIPTNGIIVVNGIDIKRKM-APDLLSCLAVSRIFLQ 59
Db 39 MAAVW---FLIVIGFPINFLIVTV-----QHKRLTPNTLLNLAVADLFW 87
Oy 60 LRTFYVNVIVIFIEFIMCSANCAILLPT---NELEW---LATWLGVEYCAKVASVR 111
Db 88 FCGFTTLYTSLHGYFVGPTGNCLEGFATLGELIALSVLAIERVYVVCCKPMNFR 147
Oy 112 ---HPL---PIW---LKMRSKLVPMWMLGSLIYVSMICVPSKYGVWVFLR--- 157
Db 148 FGENMAINGVGTWWMALACAPLVGM-----SR---IPRMOQSC 187
Oy 158 --KFSQNAATIOKEPTLAIQIFSEVAEFSVPLLEFLAVLLIFSLGRHTRPMNTVAGS 215
Db 188 GIDVYTLKPEVNE---SFVIYMFVHFTIPMIVFEFCQGLFTVKEAQAQOES-ATY 243
Oy 216 RYVGRGAPISALLSLISFLIYFSHCMIKFLSLKFRIRFILL---PIYVIGIYPS 271
Db 244 QKAEKEVTRMVIIMVIALILCWPPASVAFYIFTHQGSNESPFTMLPAFPAKASTY-- 301
Oy 272 GHSLLILGNPKLNKAMKFLLSKCC 298
Db 302 -NPVIYIMMKQF---RTCMITTLCC 323

RESULT 15
OPSD_PIG 15
ID OPSD_PIG 15 STANDARD: PRT: 348 AA.
AC 018766:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO OR RHOL.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-97475534; PubMed-9335046;
RA Petters R.M., Alexander C.A., Wells K.D., Collins E.B., Sommer J.R.,
RA Blanton M.R., Rojas G., Hao Y., Flowers W.L., Banin E.,
RA Gideclayan A.V., Jacobson S.G., Wong F.:
RT "Genetically engineered large animal model for studying cone
RT photoreceptor survival and degeneration in retinitis pigmentosa.";
RL Nat. Biotechnol. 15:965-970(1997).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.

```

Search completed: March 15, 2001, 13:20:35  
Job time: 1355 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 12:52:19 ; Search time 66.3 Seconds  
(without alignments)  
306.219 Million cell updates/sec

Title: US-09-510-332-1

Perfect score: 1521  
Sequence: 1 MEASHIITFLAVIQFLG.....GNPKLNKAKFLHKKCCQ 299

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	131	8.6	333	2	S52960	NADH dehydrogenase
2	130.5	8.6	348	1	JC4267	opsin - rabbit
3	129.5	8.5	348	1	00SH	rhodopsin - sheep
4	128.5	8.4	348	1	S23398	rhodopsin - Chinese
5	125.5	8.3	348	1	S51677	rhodopsin - rat
6	125	8.2	382	2	T26188	hypothetical prote
7	124.5	8.2	445	2	E22845	hypothetical prote
8	124	8.2	500	2	D58931	NADH dehydrogenase
9	123.5	8.1	348	1	00HU	rhodopsin - human
10	122.5	8.1	335	2	161746	phoromone receptor
11	122.5	8.1	346	2	JC5715	G protein-coupled
12	122.5	8.1	346	2	JC5715	hypothetical prote
13	121.5	8.0	348	1	A23665	opsin - mouse
14	121	8.0	355	2	I51319	RH2 opsin - green
15	121	8.0	512	1	A70201	variance factor m
16	120.5	7.9	346	2	JC5716	G protein-coupled
17	120	7.9	349	2	A45229	opsin, green-sensl
18	119.5	7.9	353	1	JN0120	rhodopsin - Japanese
19	118.5	7.8	524	2	A31318	glucose transporte
20	118	7.8	420	2	I51666	Mel-1c receptor su
21	117.5	7.7	522	2	A31556	glucose transport
22	117.5	7.7	522	2	S68362	glucose transport
23	116.5	7.7	386	1	0KFE2M	NADH dehydrogenase
24	116.5	7.7	359	2	S15403	angiotensin II rec
25	116.5	7.7	620	2	A58932	cytochrome C-type
26	115.5	7.6	348	1	00BO	rhodopsin - bovine
27	115.5	7.6	352	2	I50081	rhodopsin - green
28	115.5	7.6	355	2	A42347	opsin, green-sensl
29	115	7.6	312	1	S32696	opsin, rod - dog

30	115	7.6	538	2	E81435	probable iron-upta
31	114.5	7.5	354	1	I51200	rhodopsin - Africa
32	114	7.5	262	2	S59078	conserved hypothet
33	114	7.5	770	2	G72589	hypothetical prote
34	113.5	7.5	345	1	NMBE17	38.1k membrane pro
35	113.5	7.5	359	2	JC1104	angiotensin II rec
36	112	7.4	269	2	D70454	hypothetical prote
37	112	7.4	320	2	S20573	olfactory receptor
38	112	7.4	523	2	S06920	glucose transport
39	111.5	7.3	274	2	S42172	NADH dehydrogenase
40	111.5	7.3	354	1	S27231	rhodopsin - northe
41	111.5	7.3	355	1	A46191	iodopsin homolog
42	111.5	7.3	359	2	JC1194	angiotensin II rec
43	111.5	7.3	359	2	A42656	angiotensin II rec
44	111.5	7.3	468	2	A72619	probable NADH dehy
45	111	7.3	274	2	S42168	NADH dehydrogenase

## ALIGNMENTS

RESULT 1  
S52960 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - honeybee mitochondrion  
C:Species: mitochondrion Apis mellifera (honeybee)  
A:Variety: ligustica  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 17-Mar-2000  
C:Accession: S52960  
R:Crossier, R.H.; Crozier, Y.C.  
Genetics 133, 97-117, 1993  
A:Title: The mitochondrial genome of the honeybee Apis mellifera: complete sequence a  
A:Reference number: S52960; WUID:93114603  
A:Accession: S52960  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-333 <CHO>  
A:Cross-references: EMBL:L06178  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC4  
A:Superfamily: NADH dehydrogenase (ubiquinone) chain 2  
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidoreductase

Query Match	8.6%	Score 131;	DB 2;	Length 333;
Best Local Similarity	19.9%	Pred. No. 0.001;		
Matches	67;	Conservative	63;	Mismatches 130; Indels 76; Gaps 13;
QY	5	HLIYFLAV-----IQFLGIFTNGIIVVNGIDLKRRKAPDLLSCLAV	53	
DB	9	HMEYFLITFVLMNNSNFIOMNMEF--GTITISLIN-ISTNKPRLIYYSVI	65	
QY	54	SRFLQLEFYVNVIVPIEFIFMCSANCAILFYN---ELELW-----LATWLGVEY	103	
DB	66	SSIFL-FEMIIYLLSISTKTDTNFEMQNMFLKIGPPEHFMIVSYEMNNKQIFL	124	
QY	104	CATVAVSRHPLFIWLM-RISKLVPMILGSLIYSMLCVFHSKAGMVPFLKPFSSQ	162	
DB	125	MSLLIF-IPYVMVSMTKNSMTLYFLITNSLYIS-----FYANKFTLL	168	
QY	163	NATIQEDLAIQIFSEVAD-----FSVPLLIFFLFAVLLISGRHFRQMRNFAVGRV	217	
DB	169	KKLACSTTFNSFYFLIELNKNMFIAMIILXSPVFLISFLKNFNQNNFMYNK-	227	
QY	218	PGRCAPISALLSLISFLIYFSCHMIVELS-----SLKPHIRRFLEFLIYIGIY	269	
DB	228	-----YQMYTFLTLMFNYSMPFLSFYIKNLLIFMVSAYAMWILFLMISSM	277	
QY	270	PSGSLILLIGNPKLN-----AKKPLHS	295	
DB	278	IMTWNYIILKRVFLKMFYKNNFIDDKDKNKIYHS	313	



QY 272 GHSLLILGNPKLONAKKFLHSKCC 298  
Db 302 -NPVITIMNKKOF---RNCMLTTLCC 323

## RESULT 4

S23398

rhodopsin - Chinese hamster

C:Species: Cricetulus griseus (Chinese hamster)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S23398

R:Gale, J.M.; Tobey, R.A.; D'Anna, J.A.

J. Mol. Biol. 224, 343-358, 1992

A:Title: Localization and DNA sequence of a replication origin in the rhodopsin gene loc

A:Reference number: S23398; MID:92219256

A:Accession: S23398

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-348 &lt;GAL&gt;

A:Cross-references: EMBL:X61084; NID:949478; PIDN:CAA43398.1; PID:949479

C:Genetics:

A:Introns: 121/1; 177/2; 232/3; 312/3

C:Superfamily: vertebrate rhodopsin

C:Keywords: chromoprotein; eye; G protein-coupled receptor; glycoprotein; lipoprotein; F

F:37-61/Domain: transmembrane #status predicted &lt;TM1&gt;

F:74-96/Domain: transmembrane #status predicted &lt;TM2&gt;

F:114-133/Domain: transmembrane #status predicted &lt;TM3&gt;

F:153-175/Domain: transmembrane #status predicted &lt;TM4&gt;

F:203-230/Domain: transmembrane #status predicted &lt;TM5&gt;

F:253-276/Domain: transmembrane #status predicted &lt;TM6&gt;

F:283-309/Domain: transmembrane #status predicted &lt;TM7&gt;

F:296/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 8.4%; Score 128.5; DB 1; Length 348;  
Best Local Similarity 21.8%; Pred. No. 0.0017;

Matches 70; Conservative 63; Mismatches 129; Indels 59; Gaps 16;

QY 1 MESHLLIYFLAVIOFLGIFTNGIIVNGIDIKRHKM-APLDLISCLAVSRIFLO 59  
Db 39 MAAVM---FLIIVGFPINFLTLVTV-----QHKKLTPLNTLLNLAVADLFMV 87

QY 60 LFIYVNVIVIFIEFIFMCSANCALLFI---NELELM---LATWLGVEYCAKVASVR 111  
Db 88 FCGFTTTLTSLHGIVFEGPTGCLNLEGFATLGEIGLSLVLAIERVYVVCXPMNSFR 147

QY 112 ----HPL----FIMLKMRISKLVPMILGSLLYV--SMICVFSKYAGFMVYFLRKEFS 161  
Db 148 FGENHAIMGVAFWTMALACAP-PLVGMSTRYDEGMQSCGVY-----YTLKPEVN 199

QY 162 QNATTOKEPTLAIQIFSEVAEFSVPLILFPAVLLIFSLGRHTQMRNTVAGSRVPGC 221  
Db 200 NESFV-----IYMFVHFITPLIVIFFCYGOLVFTVKAQAQOQS-ATTQAEKE 249

QY 222 APSALLISFLILFESHQMIKVFSLKFKHRRIFL---FFILVIGIYPSGSHLL 277  
Db 250 VTRMVLIMVFFLICWFPYAGVAFYFTHOGSNFGPIFMTLPAFAFKSSIV--NPVIX 306

QY 278 ILGNPKLONAKKFLHSKCC 298  
Db 307 IMNKKOF---RNCMLTTLCC 323

RESULT 5  
S51677

rhodopsin - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S51677

R:Huber, A.; Baker, B.B.; Sander, P.; Gerdon, G.; Paulsen, R.; Williams, T.P.

Submitted to the EMBL Data Library, December 1994

A:Description: Light-history effects: Levels of rhodopsin, opsin and opsin mRNA in albin

A:Reference number: S51677

A:Accession: S51677

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-348 &lt;HOB&gt;

A:Cross-references: EMBL:Z46957; NID:9603874; PIDN:CAA87081.1; PID:9603875

C:Superfamily: vertebrate rhodopsin

C:Keywords: chromoprotein; eye; G protein-coupled receptor; lipoprotein; phosphoprote

F:296/Binding site: retinal (Lys) (covalent) #status predicted

F:322/323/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 8.3%; Score 125.5; DB 1; Length 348;  
Best Local Similarity 20.2%; Pred. No. 0.003;

Matches 66; Conservative 65; Mismatches 125; Indels 71; Gaps 16;

QY 1 MESHLLIYFLAVIOFLGIFTNGIIVNGIDIKRHKM-APLDLISCLAVSRIFLO 59  
Db 39 MAAVM---FLIIVGFPINFLTLVTV-----QHKKLTPLNTLLNLAVADLFMV 87

QY 60 LFIYVNVIVIFIEFIFMCSANCALLFI---NELELM---LATWLGVEYCAKVASVR 111  
Db 88 FCGFTTTLTSLHGIVFEGPTGCLNLEGFATLGEIGLSLVLAIERVYVVCXPMNSFR 147

QY 112 ----HPL----FIMLKMRISKLVPMILGSLLYV--SMICVFSKYAGFMVYFLR--- 157  
Db 148 FGENHAIMGVAFWTMALACAPPLVGM-----SRV-----IPEGMQSC 187

QY 158 --KFSQNTIOKEPTLAIQIFSEVAEFSVPLILFPAVLLIFSLGRHTQMRNTVAGS 215  
Db 188 GIDYTLKPEVNE---SEVYMFVHFITPLIVIFFCYGOLVFTVKAQAQOQS-ATT 243

QY 216 RVPGRGAPISALLISFLILFESHQMIKVFSLKFKHRRIFL---FFILVIGIYPS 271  
Db 244 QKAKEVTRMVIIMVIFFLICWLPYASVAMYFTHOGSNFGPIFMTLPAFAFKTASIV-- 301

QY 272 GHSLLILGNPKLONAKKFLHSKCC 298  
Db 302 -NPVITIMNKKOF---RNCMLTTLCC 323

RESULT 6  
T26188

hypothetical protein W05B5.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000

C:Accession: T26188

R:McLay, K.

Submitted to the EMBL Data Library, November 1996

A:Reference number: T20167

A:Accession: T26188

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-382 &lt;WIL&gt;

A:Cross-references: EMBL:Z82071; PIDN:CA804918.1; GSPDB:GN00019; CESP:W05B5.2

A:Experimental source: clone W05B5

C:Genetics:

A:Gene: CESP:W05B5.2

A:Map position: 1

A:Introns: 48/2; 97/3; 146/3; 185/1; 216/3; 268/2; 305/2

C:Superfamily: neurokinin 1 receptor

Query Match 8.2%; Score 125; DB 2; Length 382;  
Best Local Similarity 21.5%; Pred. No. 0.0035;

Matches 60; Conservative 54; Mismatches 105; Indels 60; Gaps 11;

QY 8 IYFLAVIOFLGIFTNGIIVV-----NGIDIKRHKMAPDLISLC 50  
Db 50 VITLASM--VIGVGNITLVVVVATNKSVMGRMGCTGRKNRVVCSMPKRNALMLVLNM 107

QY 51 LAVSRIFQLFTFYVNVIVIFIEFIFMCSANCALLFIENELEMLATVGLVY-CAKVAS 109  
Db 108 LMIADLLIFLPLVVDVTKTFWFSVAFCKSVFNNTSVVYSIMSLVFTICERRA 167





A:Residues: 248-263 <ING>  
R:Sheffield, V.C.: Fishman, G.A.; Beck, J.S.; Kimura, A.E.; Stone, E.M.  
Am. J. Hum. Genet. 49, 699-706, 1991  
A:Title: Identification of novel rhodopsin mutations associated with retinitis pigmentosa  
A:Reference number: 151864; MUID:91377732  
A:Accession: 151864  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 14-16, 'M', 18-20 <SHE>  
A:Cross-references: GB:S55797; NID:g235658; PIDN:AAB19830.1; PID:g235659  
A:Accession: 164813  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 179-181, 'S', 183-185 <SH2>  
A:Cross-references: GB:S55799; NID:g235660; PIDN:AAB19831.1; PID:g235661  
A:Accession: 164814  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 264-266, 'L', 268-270 <SH3>  
A:Cross-references: GB:S55874; NID:g235662; PIDN:AAB19832.1; PID:g235663  
C:Genetics:  
A:Gene: GDB:RHO  
A:Cross-references: GDB:120347; OMIM:180380  
A:Map position: 3q21.3-3q24  
A:Introns: 121/1; 177/2; 232/3; 312/3  
A>Note: defects in this gene can result in retinitis pigmentosa  
C:Superfamily: Vertebrate rhodopsin  
C:Keywords: acetylated amino end; chromoprotein; eye; G protein-coupled receptor; glycoprotein

F:1-36/Domain: extracellular #status predicted <EX1>  
F:37-61/Domain: transmembrane #status predicted <TM1>  
F:62-73/Domain: intracellular #status predicted <IN1>  
F:74-99/Domain: transmembrane #status predicted <TM2>  
F:100-113/Domain: extracellular #status predicted <EX2>  
F:114-133/Domain: transmembrane #status predicted <TM3>  
F:134-152/Domain: intracellular #status predicted <IN2>  
F:153-175/Domain: transmembrane #status predicted <TM4>  
F:176-202/Domain: extracellular #status predicted <EX3>  
F:203-230/Domain: transmembrane #status predicted <TM5>  
F:231-252/Domain: intracellular #status predicted <IN3>  
F:253-276/Domain: transmembrane #status predicted <TM6>  
F:277-284/Domain: extracellular #status predicted <EX4>  
F:285-309/Domain: transmembrane #status predicted <TM7>  
F:310-348/Domain: intracellular #status predicted <IN4>  
F:1/Modified site: acetylated amino end (Met) #status predicted  
F:2,15/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:110-187/Disulfide bonds: #status predicted  
F:266/Binding site: retinal (Lys) (covalent) #status predicted  
F:322,323/Binding site: palmitate (Cys) (covalent) #status predicted  
F:334,338,343/Binding site: phosphate (Ser) (covalent) (by rhodopsin kinase) #status predicted  
F:336/Binding site: phosphate (Thr) (covalent) (by rhodopsin kinase) #status predicted

Query Match 8.1%; Score 123.5; DB 1; Length 348;  
Best Local Similarity 20.8%; Pred. No. 0.0044;  
Matches 68; Conservative 63; Mismatches 125; Indels 71; Gaps 16;

QY 1 MLESHLITFLAVIYOFELGIFGTNGIIVVNGIDILHRKM-APLDLISCLAVSRIFLQ 59  
DB 39 MIAAYM---FLIIVLGPPEINFLTYIV-----OHKRLRPPLNLIINLAVADLFMY 87  
QY 60 LEIFVAVYIVIFEEIFMCSANCAILLFI-----NELEIM---LATWLGVGYCAKVASVR 111  
DB 88 LGGLSTLYTSLHGFTVGCNLEGFATLGEIAMSILVLAIRYVVVCCPMNFR 147  
QY 112 ---HPL---FIW---LKMRIKIVPMIIGSLIYVSMICVFRFSKXAGVNFYFLR--- 157  
DB 148 FGENNAINGVAFVTVMALACAAPLAGV-----SRV---IPGLDSC 187  
QY 158 --KFSNANTIOKEPTLAIQIFSEVAEFSVPLIFLFAVLLIFSLGHRTRQMRNTVAGS 215  
DB 188 GIDYITLAKPEVANE---SFVIYMFVHTITMIIIFCYGGLVFTVKRPAANOQDS-ATT 243

```

OY      216 RVPGRGAPISMLSLPSLLYLFEHSCKMKVFLSKLFFIRFIEP----FTLLVIQIYP 2/1
Db      244 QKAEKEVTNRMMIVIMIAVLICWPIYSAAFIETFHQSNGFGIFMTITAFPAFKSAAYT-- 301
OY      272 GHSLLILGNPKLKONAKRFLLSKCC 298
           :| |: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      302 -NPVIYIMANKKF-----INCMLTITCC 323

RESULT 10
161746
pheromone receptor VN4 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 29-Sep-1999
C:Accession: 161746
R:Dulac, C.; Axel, R.
Cell 83, 195-206, 1995
A>Title: A novel family of genes encoding putative pheromone receptors in mammals
A:Reference number: A57223; MUID:96028094
A:Accession: 161746
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-335 <RES>
A:Cross-references: EMBL:U36896; NID:g1055249; PIDN:AAC52285.1; PID:g1055250
C:Superfamily: pheromone receptor VN3t

```

[illegible]



RESULT 14  
 RH2 opsin - green anole  
 C:Species: Anolis carolinensis (green anole)  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 13-Aug-1999  
 C:Accession: U51319  
 R:Kawamura, S.; Yokoyama, S.  
 J. Mol. Evol. 40, 594-600, 1995  
 A:Title: Paralogous origin of the rhodopsin-like opsin genes in lizards.  
 A:Reference number: U51319; MUID:95371134  
 A:Accession: U51319  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-355 <RAM>  
 A:Cross-references: GB:S79167, NID:q1042072; PIDN:AAB35062.1; PID:q1042073  
 C:Genetics:  
 A:Gene: rh2AC  
 A:Introns: 121/1; 177/2; 232/3; 312/3  
 C:Superfamily: vertebrate rhodopsin  
 C:Keywords: chromoprotein; eye; G protein-coupled receptor; lipoprotein; phosphoprotein;  
 F:296/Binding site: retinal (lys) (covalent) #status predicted  
 F:322,323/Binding site: palmitate (cys) (covalent) #status predicted

Query Match 8.08; Score 121; DB 2; Length 355;  
 Best Local Similarity 23.68; Pred. No. 0.0071;

Matches 62; Conservative 43; Mismatches 112; Indels 46; Gaps 12;

```

OY 8 IYFLAVIQFLIGFTNGIIVVNGIDLIKRRK-APDLLSCLAVSRIFLOLFITYN 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 44 IFFL-----IFTGPIRILFLVTFKHKLRPLNVLVMAVDLFMACFEYVT 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 67 VIVIFEFIFMCSANCAILLFT----NELW----LATWLGVCYCAKVASVR----HPL 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95 FYTANGFYIFGPIGCALEGFAILGGOVALMSLVLAIERIYVCKPMGNFRSATHAL 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 115 ----FTWLMKRIKSLVPMWILGSLYV--SMICVFHSKYAGFMPYFLRKFFSONATIQK 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 155 MGISTWF-MRSCSACP-PLIGMSKRIPEGMCSCGPDY-----YTLNPDYHN 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 169 EDTLAIQIFSEVAEFSVPLILFLPAVLLILFSLGRHTRQNTVAGSRVPGAPISALL 228
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 201 ESYV---LYMEGVHVFIPVVIFFFSYGRILCKVREAAQOQES-ASTOKAREVTRMYIL 256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 229 STLSFLILFESHCMIKVFLSLK 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 257 MYLGFLLMTPTAMVAVFWLFTNK 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

# RESULT 15

A70201  
 virulence factor mv1N protein homolog precursor - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Dec-1999

C:Accession: A70201  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Ladhigra, R.; White  
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943

A:Accession: A70201  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-512 <KLE>  
 A:Cross-references: GB:AE001179; GB:AE000783; NID:g2688738; PIDN:AAC67146.1; PID:g268874

A:Experimental source: strain B31

C:Superfamily: mv1N protein  
 F:1-28/Domain: signal sequence #status predicted <SIG>  
 F:29-512/Product: virulence factor mv1N protein homolog #status predicted <MAT>

Query Match 8.08; Score 121; DB 1; Length 512;  
 Best Local Similarity 21.24; Pred. No. 0.0097;  
 Matches 60; Conservative 43; Mismatches 74; Indels 106; Gaps 13;

```

OY 5 HLIIYFLAVIQFLIGFTNGIIVVNGIDLIKRRKAPDLLSCLAVSRIFLOLFITY 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 YLVYIILILSLSI-----FVSLNSYKRIFFIPSPIMLSFG-----ILISFLFY 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 65 -----VWVIVIFEFIFMCSANCAILLFELMLATWLGVCYCAKVASVRHPLF 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 183 GREGISAVIGVIRGGFLQFLIPFANCLMIGF-----AMKPTFY-----FREKVF 227
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 116 I-----WLKM-----RISKLVPMWILGSLVYSMICVFHSKYAGFMPYFLRK 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 228 INFILTRWLMIFGFSISITTOIS-----PALASTLEIGSVISLSNAVYQLPVGI--F 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 160 EQONATIQKEDTLAIOIFSEVAEFSV-----PLILFLPAVLLILFSLGRHTRQNT 211
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 281 YISIAIV-----ITPKMAEHAVLGNNIKLNALVDGIRKILLIF----- 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 212 VAGSRVPGRGAPISALLSTLSFLILFESHCMIKVFLSLKRI 254
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 320 -----IP-----VSFLMFIWSDYILNLFILMGKFSI 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: March 15, 2001, 13:00:06  
 Job time: 467 sec





```

484 .....CAGCACAGAGAGCTGCGACCGCTCTCAATACATCTGTC 525
50 CysleuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyrValAs 66
|||||
526 AACCTAGCGGTGGCTGACCTTCATAGTCTCTATGAGTGGCTTACACGAC 575
66 nValIleValIlePhePheIleGluPheIleMetCysSerAlaAsnGlySA 83
|||||
576 CCTCTACACCTCTCTGATGATGACTTCGCTTCCGGGCCACAGATGCA 625
83 lAlleLeuLeuPheIle.....AsnGluLeuLeuTyr... 94
|||||
626 ATTGGAGCGCTTCTTGGACCCCTGGCGGAAATTCCTGCTGTC 675
95 .....LeuAlaThrTyrPheGluGlyValPheTyrCysAlaIleValAl 108
|||||
676 TTGCTGCTCTGCGCATCGAGCGGTACGGTGTGTGTGTAACCCATGAG 725
108 aserValArg.....HisProLeu.....PheIleTyr 117
|||||
726 CAACCTTCCCTTCGGGAGAACCATCCATCATGGCGCTTGCCCTTACCT 775
117 TP.....LeuLysMetArgIleSerLysLeuValProThrMetIle 130
|||||
776 GGGTCATGGCGCTGGCTGGCGGACCCCACTCGCGCTG..... 819
131 LeuGlySerLeuLeuTyrValSerMetIleCysValPheHisSerLysTyr 147
|||||
820 .....TTCAGGTA 827
147 fAlaGlyPheMetValProTyrPheLeuArg.....LysP 159
|||||
828 C.....ATCCCGAGCGCTGAGTCTGCTGTGGAATGACT 865
159 hepPheSerGlnAsnAlaThrIleGlnLysGluAspThrLeuAlaIleGln 175
|||||
866 ACTACACGCTCAAGCGGAGTCAACAAGAG.....TCTTTTTC 906
176 lIlePheSerPheValAlaGluPheSerValProLeuLeuIlePheLeuPh 192
|||||
907 ATCTACATGTTCTGTGTCACATTCATCCCATGATATCATCTCTTTT 956
192 eAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArgIleMetA 209
|||||
957 CTGCTATGGGCGCTGCTCTTCAACGCTCAAGAGCGCTCCACGACG 1006
209 gAsnThrValAlaGlySerArgValProGlyArgGlyAlaProIleSer 225
|||||
1007 AGGAGTCA...GCCACACACAGAGAGGACAGAGAGGCTCACCGCATG 1053
226 AlAlaLeuSerIleLeuSerPheLeuIleLeuTyrPheSerHisCysMe 242
|||||
1054 GTCATCATCATGTCATGCTCTTCTATGTCGTGGGCGCTTACGCCGAG 1103
242 tIleLeuValPheLeuSerSerLeuLysPheHisIleArgArgPheIleP 259
|||||
1104 CGTGGCATTTACATCTTCAACGCGCTCCAACTCGGTCGCATCT 1153
259 heLeu.....PhePheIleLeuValIleGlyIleTyrProSer 271
|||||
1154 TCATGACCATCCACGCTTCTTGGCAAGAGCGCGCATCTAC..... 1197
272 GlyHisSerIleLeuIleLeuGlyAsnProLysLeuLysGlnAsnAl 288
|||||
1198 ...AACCGTGCATCTATATCATGATGACACAGCATTC..... 1233
288 AlLysLysPheLeuLeuHisSerLysCysGly 299
|||||
1234 .CGGACATGCTGCTCACACCATCTGCTGCGC 1266
seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-033-081B-1
seq_documentation_block:

```

```

; Sequence 1, Application US/08033081B
; Patent No. 5498521
; GENERAL INFORMATION:
; APPLICANT: Dryja, Thadeus P.
; APPLICANT: Berson, Eliot L.
; TITLE OF INVENTION: DIAGNOSIS OF HEREDITARY RETINAL
; TITLE OF INVENTION: DEGENERATIVE DISEASES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; COMPUTER: IBM PS/2 Model 50z or 55sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/033,081B
; FILING DATE: March 11, 1993
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/825,296
; FILING DATE: January 23, 1992
; APPLICATION NUMBER: 07/469,215
; FILING DATE: January 24, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00246/069005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3016
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-033-081B-1

alignment_scores:
      Quality: 124.50      Length: 328
      Ratio: 0.759      Gaps: 16
      Percent Similarity: 50.000      Percent Identity: 20.732

alignment_block:
US-09-510-332-1 x US-08-033-081B-1 ..

Align seg 1/1 to: US-08-033-081B-1 from: 1 to: 3016

1 MetLeuGluSerHisLeuIleIleTyrPheLeuLeuAlaValIleGlnPh 17
|||||
409 ATGCTGCGCGCTACATG.....TTTCTGTCATGTCGTCTGGGCTT 449
17 eLeuLeuGlyIlePheThrAsnGlyIleIleValValAlaAsnGlyIleA 34
|||||
450 CCCCATCACTTCTCTACGCTCTACGTCACCGCT..... 483
34 sPheUleLysHisArgLysMet...AlaProLeuAspLeuLeuSer 49
|||||
484 .....CAGCACAGAGAGCTGCGACCGCTCTCAATACATCTGTC 525
50 CysLeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyrValAs 66
|||||
526 AACCTAGCGGTGGCTGACCTTCATAGTGTCTTACAGTGGCTTACACGAC 575
66 nValIleValIlePhePheIleGluPheIleMetCysSerAlaAsnGlySA 83

```

[illegible]

```

725 ..TGGGCGCTTTTG...TCCGTTTGTGTCCTGTTTCTGCGCTT 681
144 HIserytYrAlaGlyRheMetValProTyRhe..... 155
680 CCGGCTTCTTTC.....TTTTTGTCCTTTTTCGTTGTCGCTC 637
156 .....LeuArgLysRhePheSerGlnAsnAlaT 165
636 CCGCTTCTCTGTTGTGTTGCTGCTGCTTTTTCCTTCCTTC 587
165 hrlleGlnLysGlnAspThrIleAlaIleGlnIlePheSerPheValA 181
586 CTGTG.....GTGGCTTTTGTGCTTTTTCCTTTTCGCGG 552
182 GluPheSerValProLeuLeuIlePheLeuPheAlaValLeuLeu 198
551 CCGTTTGGGTTCCCTTTTGTGCTTTCGCTTTCGCTTTCCTTCT 502
198 eRheSerLeuGlyArgHisThrArgGlnMetArgAsnThrVal...AlaG 214
501 TTTCTGCGTTTCTTTTTCGTTTTCGCTTTCCTTTCTTTTTC 452
214 LysArgValProGlyArgGlyAlaProIle.....SerAlaLeu 228
451 TCTCGTCCCGCTTTCCTTTCGCTGCTGCTTTTTCCTTTTCGT 402
229 SerIleLeuSerPheLeuIleLeuTyRheSerHisCysMetIleLys 245
401 TGTCTCTGTCGTTGTCCTTTTTCCTTTTTCCT.....TTCGCT 361
245 lPheLeuSerSerLeuLysRheHisIleArgArgPhe...IlePheLeu 261
360 CTTTCTG...CGGCTTGGCTTTCGCGCTTTTTCCTTTTCGCTTTC 314
261 hePheIleLeuVal 265
313 TTTTTCGCTGCT 300

seq_name: /cgn2_6/ptodata/2/lna/5b_COMB.seq:US-08-820-521-1

seq_documentation_block:
: Sequence 1, Application US/08820521
: Patent No. 5942416
: GENERAL INFORMATION:
: APPLICANT: Bergsma, Derk
: APPLICANT: Ganesh, Sathe
: APPLICANT: Fuelterer, Wendy
: APPLICANT: Mao, Joyce
: TITLE OF INVENTION: CDNA CLONE HNFY20 THAT ENCODES
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/820,521
: FILING DATE: 19-MAR-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Han, William T
: REGISTRATION NUMBER: 34,344

```

```

: REFERENCE/DOCKET NUMBER: GH50011
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-5219
: TELEFAX: 610-270-4026
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1841 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-820-521-1

alignment_scores:
: Quality: 122.50 Length: 296
: Ratio: 0.833 Gaps: 18
: Percent Similarity: 49.662 Percent Identity: 23.986

alignment_block:
US-09-510-332-1 x US-08-820-521-1 ..
Align seg 1/1 to: US-08-820-521-1 from: 1 to: 1841

10 PheLeuLeuAlaValIleGlnPheLeuLeuGlyIlePheThrAsnGlyI 26
341 TTTCTGCGTACCTTCCTCCTTCCTGTTGGGCTCCCTCAAC...CT 387
26 eileValAlaValAsnGlyIleAspLeuIleLysHisArgLysMetAla 43
388 GCTGGCCCTGTGTCTTCGTTGGCAAGCTGCACGCCGCCGCTGCC. 436
43 roleAspLeuLeuLeuSerCysLeuAlaValSerArgIlePheLeuGln 59
437 ..GTGGACGTGCTCTCTCACTGACCCGCTGACGCTGCTGCTG 484
60 LeuPheIlePheTyValAsnValIleValIle.....PheP 72
485 CTGTCTCTGCTTTCCTCCATGCTGGAGCAGCCAAATGCATGCATGCC 534
72 eileGlnPheIleMetCysSerAlaAsnCysAlaIleLeuLeuPheIle 89
535 CCGCCCTTCATCTCTGCGCCACTCTGGA.....TTCATCTTCTTCA 578
89 snGluLeuGluLeuTyRProLeuAlaThrIlePheGlyValPheTyCysAla 105
579 CCACCATCTATCTC...ACGCGCTCTCTCTGCGACGCTGAGCATTTGA 625
106 LysValAlaSerValArgHisProLeuPheIleTyRLeuLysMetArg. 121
626 CGCTTCCTGAGTGTGGCCACGCCCTG.....TGTACAAAGCCGCCG 669
122 .....IleSerLysLeuValProTyRMetIleLeuG 132
670 GAGGCTGGGCGACGAGTCTGTGAGTGTGGCTGCTGCTGCTGCT 719
132 Ly.....SerLeuTyRValSerMetIleCys.....ValPhe 143
720 CTGCTACATGCACGCTGCTCTACGCTATAGAAATTCACAGGACATCTCC 769
144 HIserytYrAlaGlyRheMetValProTyRheLeuArgLysPheP 160
770 CACAGCCAG..... 778
160 eSerGlnAsnAlaThrIleGln.....LysGlnAspThrLeuAlaI 174
779 ..GCGACCAATGGACCTGCTACCTGAGTTCGGAAGCACGCTAGCA 827
174 le.....GlnIlePheSerPheValAlaGlyPheSerValPro 186
828 TCTCTCTGCCCTGGGCGTGGAGATGCTGTGCTCTTTCGTGCTCCG 877

```



```

187 leuileupheleupheleuAlaValleuLeuileupheserleuGlyIar 203
    |||:|||||:..... |||:|||||:
878 CTGATCATCACCACGACTGCTGTACAGCCGCTGTGTGATGATCCCTGGCAG 927
203 G.....HisThrArgInMetaArgAsnThrValAlaGlySerArgV 217
    |||:|||||:..... |||:|||||:
928 AGGGGCGACCGCCGCGCAGAGAG..... 955
217 aLpGgGlyArgGlyAlaProlleSerAlaLeu.....Serleu 231
    |||:|||||:..... |||:|||||:
956 .....GTGGGGGGCTGTGGCGCGCCACGCTGCTC 985
232 SerPheleuileuLeuTYR.....PheSerHis..... 240
    |||:|||||:..... |||:|||||:
986 AACTTCCTTGTCTGCTTGTGGCCCTACACAGTGTCCATGTGTGGGCTA 1035
241 ...Cys.....MetleuValPheleuSerS 249
    |||:|||||:..... |||:|||||:
1036 TATCTGCGGTGAAGCCGCGCTGGAGATCTACGTGACGCTTCACGA 1085
249 erleuLeupheHisIleArgArgPheleuPhe 261
    |||:|||||:..... |||:|||||:
1086 CCTGAACCTCTGTGTGACGCCCTTGTCTACTCTC 1123
seq_name: /cgn2_6/ptodata/2/lna/6_COMB.seq:US-08-728-603-16

seq_documentation_block:
; Sequence 16, Application US/08728603
; Patent No. 6093806
; GENERAL INFORMATION:
; APPLICANT: Cesarman, Ethel
; APPLICANT: Knowles, Daniel M.
; TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,603
; FILING DATE: 10-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BRAMAN, SUSAN J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 19603/720
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1636
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1202 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-728-603-16

alignment_scores:
Quality: 120.50 Length: 263
Ratio: 0.867 Gaps: 14
Percent Similarity: 52.852 Percent Identity: 22.433

```

```

alignment_block:
US-09-510-332-1 x US-08-728-603-16 ..
Align seg 1/1 to: US-08-728-603-16 from: 1 to: 1202

11 leuLeuAlaValIleGlnPheleuLeuGlyIlePheThrAsnGlyIleI 27
    |||:|||||:..... |||:|||||:
203 ATACTCTCTGAT...TTCTCATTAATAGTCTTGGAATAGATTTGGG 249
27 eValValAlaValAsnGlyIleAspleuileuYHisArgIysMeta...P 43
    |||:|||||:..... |||:|||||:
250 CACTACAT...TTTTCACAGCACCGATCCGCGGAGAG 287
43 roleuAspleuLeuLeuSerCysleuAlaValSerArgIlePheleuGln 59
    |||:|||||:..... |||:|||||:
288 CGATAGATATACCTCTCTCGGTATCTGCTTAACCTGCTGTCTTACG 337
60 leuPheIlePheTYrValAlaValIleValIlePhePheIleGlnPheI 76
    |||:|||||:..... |||:|||||:
338 ATATCTTA...TTGGCAGAGCTGTGATGTTTGTTCATAT 381
76 eMetCysSerAlaAsnGlyAlaIleleuLeupheIleAsnGlyIleuGlu 93
    |||:|||||:..... |||:|||||:
382 CATCTCCACAGCTGTGTGACAGACTGAATTTT...T 419
93 eutPleuAlaIleThrIleuGlyValPheTYrCysAlaIysValAlaSer 109
    |||:|||||:..... |||:|||||:
420 ACTATTATATGATCTACTTGTGATATCTTACAGTGTGTGTGCTGCTA 469
110 ValArgHisProlleuPheIle.....TripleuYsMetar 121
    |||:|||||:..... |||:|||||:
470 GTGAGTACCTCTGCTGTGCGCATATCTACGCGTTCCTCGCCCAAG.... 514
121 gIleSerIysleuValProtrPmetIleleuGlySerleuLeuTYrValS 138
    |||:|||||:..... |||:|||||:
515 ...AAGCAGTCCCTCGGATGGTACATGACATCGCTCAGCTTAATG 560
138 eMetIle.....CysValPheHisSerIysTYrAlaGly 149
    |||:|||||:..... |||:|||||:
561 CATGTGCTGTGCGGGGATGCTGTGCACACAGAGAGGAGGTGTCGAC 610
150 PheMetValProTYrPheleuArgIysPhePheSerGlnAsnAlaThrI 166
    |||:|||||:..... |||:|||||:
611 .....CCGTCAGCAGACGCGCATGTGTATGGAAGCCGGA 651
166 eGlnIysGluAsp...ThrleuAlaIleGlnIlePheSerPheValAlaG 182
    |||:|||||:..... |||:|||||:
652 CATGACTGACAGCTGGCAGCTGCATGTCAAGACCGTTCATCTGACAG 701
182 lUpheSerValProlleuLeuIlePheleuPheAlaValleuLeuIle 198
    |||:|||||:..... |||:|||||:
702 GTTTCCTGTATACCCCTG.....GCCCTCTTATTCGTGTT 736
199 PheSerleuGly.....ArgHisThrArg...GlnMetArgAs 210
    |||:|||||:..... |||:|||||:
737 TATGCTCTACCTGTGTGTGTGTGAGAGACAAACCTGACCTGACAGG 786
210 nThrValAlaGlySerArgValProlGlyAlaProlleSerAla 227
    |||:|||||:..... |||:|||||:
787 GAAGGTAAGGGG..... 799
227 euleuSerIleLeuSerPheleuIleleuTYrPheSerHisCysMetIle 243
    |||:|||||:..... |||:|||||:
800 .....GTGATTGTGCTGTGTGCTGTGTTTGTGTTTGTGTTTCCCT 844
244 LysValPheleuSerleuYsPheHisIleArgArg 256
    |||:|||||:..... |||:|||||:
845 TACCACGTAATAATCTACTGAGACACTCTGTAAGGCA 883
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-770-379-20

seq_documentation_block:
; Sequence 20, Application US/08770379

```

```

: Patent No. 5849564
: GENERAL INFORMATION:
: APPLICANT: Chang, Yuan
: APPLICANT: Bohenzky, Roy A.
: APPLICANT: Russo, James J.
: APPLICANT: Edelman, Isidore S.
: APPLICANT: Moore, Patrick S.
: TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
: TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESS: Cooper & Dunham LLP
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/770,379
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 52342
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 32207 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-770-379-20

```

```

alignment_scores:
  Quality: 120.50      Length: 263
  Ratio: 0.867        Gaps: 14
  Percent Similarity: 52.852  Percent Identity: 22.433

```

```

alignment_block:
  US-09-510-332-1 x US-08-770-379-20 ..

```

```

Align seg 1/1 to: US-08-770-379-20 from: 1 to: 32207

```

```

11 LeuenuAlaValIleGlnPheLeuLeuGlyLePheThrSnglyIleI 27
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24227 ATATCTCTCTGATT...TTCTCTATAATGCTTTGTAATGATGCT 24273
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27 eValValAlaValIleAspLeuLeuIleYSHsArgLysMetAla...P 43
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24274 CACCTACAT...TTTTCAMACACCGATCCGGCGGACAGAG 24311
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 rOlAuAspLeuLeuLeuSerCysLeuAlaValSerArgIlePheLeuGln 59
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24312 CGATAGATATACCTCTCTGCTGATTCCTTAACCTGCTGCTTAGC 24361
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 LeuPheIlePheTyValAsnValIleValIlePhePheIleGlnPheI 76
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24362 ATATCTCTA...TTGGCAGAAAGTGTATGTTGTTGTTCCCAATAT 24405
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 eMetCysSerAlaAsnCysAlaIleLeuLeuPheIleAsnGluLeuGlu 93
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24406 CACTCCACAGAGGCTTGTCAGACTTGAAATTTTTTTT...T 24443

```

```

93 euTriPheuAlaThrTriPheuGlyValPheTyrcysAlaLysValAlaSer 109
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24444 ACTATTATATCTCTACTGATATCTCAGTGTGCTGCTGCTGCTA 24493
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 ValArgHisProLeuPheLeu.....TriPheuLysMetArg 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24494 GTGAGGTACTCTCTGTCGATATTTACGGCTTCTGCCCCAG... 24538
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 gIleSerLysLeuValProTriPheIleLeuGlySerLeuLeuTyValS 138
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24539 ...AAGCAGTCCCTCGATGAGTACGATCCGCTGCTGCTGCTGCT 24584
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 emetIle.....CysValPheHisSerLysTyValAlaGly 149
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24585 CATTGCTGCTCTCGGGGATGCCCTGTCGACAGAGAGGAGGCTGCAC 24634
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 PheMetValProTyrrPheLeuArgLysPhePheSerGlnAsnAlaThrI 166
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24635 .....CCGGTCAGACAGACAGCCCATGTTATGAAACGGCGGAA 24675
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 eGlnLysGluAsp...ThrLeuAlaIleGlnIlePheSerPheValAlaG 182
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24676 CATGACTGCAGACTGCGGAGCTGATGTCAGAACCGTGTCACTGTCAG 24725
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
182 LysPheSerValProLeuLeuIlePheLeuPheAlaValLeuLeuLeu 198
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24726 GTTTCCTCTTACCCCTG.....GCCCTCTTATCTGTGTT 24760
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 PheSerLeuGly.....ArgHisThrArg...GlnMetArgas 210
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24761 TATCTCTCAGCTGCTGCTGCTGAGAGAGCAAAAGCTGCAAGCAGCG 24810
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 nTriValAlaGlySerArgValProGlyArgGlyAlaProIleSerAla 227
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24811 GAAGGTAAAGGGG..... 24823
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
227 euLeuSerIleLeuSerPheLeuLeuLeuTyrrPheSerHisCysMetIle 243
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24824 ....GTGATTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 24868
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 LysValPheLeuSerSerLeuLysPheHisIleArgArg 256
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24869 TACCAGCTACTAATCTAGTACACTCTGCTAAGCGCA 24907
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-466-103A-1
seq_documentation_block:
: Sequence 1, Application US/08466103A
: Patent No. 5856124
: GENERAL INFORMATION:
: APPLICANT: Reperet, Steven M.
: APPLICANT: Ebisawa, Takashi
: TITLE OF INVENTION: HIGH-AFFINITY MELANOTONIN
: TITLE OF INVENTION: RECEPTORS AND USES THEREOF
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESS: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/466,103A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/319,887

```

```

? FILING DATE: 07-OCT-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/261,857
? FILING DATE: 17-JUN-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Fraser, Janis K.
? REGISTRATION NUMBER: 34,819
? REFERENCE/DOCKET NUMBER: 00786/250002
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 617/542-5070
? TELEFAX: 617/542-8906
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1320 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: Coding Sequence
? LOCATION: 32...1291
? OTHER INFORMATION:
?
? US-08-466-103A-1

alignment_scores:
? Quality: 118.00 Length: 294
? Ratio: 0.752 Caps: 12
? Percent Similarity: 53.401 Percent Identity: 19.388

alignment_block:
US-09-510-332-1 x US-08-466-103A-1 ..
Align seg 1/1 to: US-08-466-103A-1 from: 1 to: 1320

2 LeuGluserHisLeuLeuLeuLeuLeuLeuValIleGlnPhe 18
111 111 111 : : : : : : : : : : : : : : : : : :
122 CTGACCTCCCTCGCGCGGTCTTATATTCACCATGTTGTGATGT 171
18 uLeuGlyIlePheThrAsnGlyIleIleValValValAsnGlyIleAsp 35
: : : : : : : : : : : : : : : : : : : : : : : :
172 CTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 200
35 euLeuLeuHisArgIleMetAlaProLeu...AspLeuLeuLeuSerCys 50
: : : : : : : : : : : : : : : : : : : : : : : :
201 TCCGAGAGAACAAAGACGCTGCAGATCTGGAATCTTGTGTGTCAGT 250
51 LeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyValAsnVa 67
: : : : : : : : : : : : : : : : : : : : : : : :
251 TTGCTCTATGGCGATCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
67 IleValIlePhePheIleGlnPheIleMetCysSerAlaAsnGlyAla 84
: : : : : : : : : : : : : : : : : : : : : : : :
301 CATAGCTATTTCAGATGATGATGATGATGATGATGATGATGATGATG 350
84 LeuLeuPheIleAsnGlnLeuGlnLeuTrpLeuAlaIleThrPheGly 100
: : : : : : : : : : : : : : : : : : : : : : : :
351 TCAGTGGCTCTCATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 385
101 ValPheTyCysAlaValValAlaSerValArgHisProLeuPheIleTr 117
: : : : : : : : : : : : : : : : : : : : : : : :
386 TCAGTCTTCACATACAGCCATGACTTATCAACAGCTTGTGCTCATCTG 435
117 PheuLysMetArgIleSerLysLeuVal.....ProTrpMetI 130
: : : : : : : : : : : : : : : : : : : : : : : :
436 CCACAGCCCTGAGATATGACAAAGCTTATTAATCAAGAACAGCCCTG 485
130 LeuGlySerLeuLeuTyValSerMetIleCysValPheHisSerLys 146
: : : : : : : : : : : : : : : : : : : : : : : :
486 ACCCTGGCTGACATGATGATGATGATGATGATGATGATGATGATGAT 535
147 TyrAlaGlyPheMetVal.....ProTyPheLeuArgLysPhePhe 161
: : : : : : : : : : : : : : : : : : : : : : : :

```

```

536 TTTGTTGATCTACAGTATGACCCCAAGATTTTCTGTGCATTTGC 585
161 rGlnAsnAlaThrIleGlnLysGlnAspThrLeuAlaIleGlnIlePhe 178
: : : : : : : : : : : : : : : : : : : : : : : :
586 GCAG.....ACAGTGAATCTCTATACACCATTAACAGTATGATGTCG 629
178 erPheValAlaGlnPheSerValProLeuLeuIlePhePheAlaVal 194
: : : : : : : : : : : : : : : : : : : : : : : :
630 ATTTTATGATCCCTCTAGTGTGTGACATTCCTGTTACTTA...AGAATA 676
195 LeuLeuLeuIlePheSerLeuGlnArgHisThrArgGln..... 207
677 TGGCTTTAGTATGATCAAGTCAAAACAGAGTATGACAACTTCAAGCA 726
208 .....MetArgAsnThrValAla..... 213
727 AAAGTGCACAAACAGACTTGAGAAATTTCTTGACCATGTTGTGCTCT 776
214 .....GlySerArg 216
777 TTGTACTTTTGCAGTTGCTGGGCCCCCTTAACCTTATCGGCTTGTCT 826
217 Val.....ProGlyArgGlyAlaPro...IleSerAlaLeuLeu 229
827 GTGGCATTAATCCGTTTCATGTGGCACCAAGATTCACAGATGCGTGT 876
229 rIleLeuSerPheLeuLeuLeuTyPheSerHisCysMetIleLysValP 246
877 TGTTTTAAGCTATTTCAATGCGCTTATTTTAACAGTTGCTCAATGCTCT 926
246 heLeuSerSerLeuLysPheHisIleArgArg 256
927 TATATGCTGCTTAATCAAAACCTCCGCAAG 958

seq_name: /cgn2/6/ptodata/2/ina/5B_COMB.seq:US-08-446-855A-1

seq_documentation_block:
? Sequence 1, Application US/08446855A
? Patent No. 5849573
? GENERAL INFORMATION:
? APPLICANT: Stewart, Thomas S
? APPLICANT: Flores, Maria V
? APPLICANT: O'Sullivan, William J
? TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
? TITLE OF INVENTION: phosphate synthetase II
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Nixon & Vanderhye PC
? STREET: 1100 No. 5849573th Glebe Road, 8th Floor
? CITY: Arlington
? STATE: Virginia
? COUNTRY: USA
? ZIP: 22201-4714
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.24
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/446,855A
? FILING DATE: 06-Jul-1995
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Mitchard, Leonard C
? REGISTRATION NUMBER: 29,009
? REFERENCE/DOCKET NUMBER: 47-80
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 703-816-4100
? TELEFAX: 703-816-4100
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 8920 base pairs
? TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
US-08-446-855A-1

alignment_scores:
    Quality: 106.00      Length: 350
    Ratio: 0.697         Gaps: 17
    Percent Similarity: 43.429    Percent Identity: 20.857

alignment_block:
US-09-510-332-1 x US-08-446-855A-1/rev ..

Align seg 1/1 to reverse of: US-08-446-855A-1 from: 1 to: 8920

4 SerHisLeuIle..IleTyrPheLeuLeuAlaIleGlnPheLeuLeuG 20
|||||
6560 AGTCATCTTCTTCATCTTCATCATTAATTAATCATCATCATCATCA 6511
20 Lyle...PheThrAsnGlyIleIleValAlaValAsnGlyIleAspLeu 35
|||
6510 GTATTCATTAATCTTCACTACTAATAATTAATTAATTAATTCCTCAT 6461
36 IleLysHisArgLysMetAlaProLeuAspLeuLeuLeuSerCysLeuAl 52
|||
6460 GTTACCTTGTGACGAATG.....TCATAAAT 6435
52 aValSerArgIlePheLeuGlnLeu.....PheIlePheTyrAla 66
|||
6434 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6385
66 sNValIle.....ValIlePhePheIle 73
|||
6384 ACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6335
74 GluPheIleMetCysSerAlaAsnCysAlaIleLeuLeuPheIleAsnG 90
|||
6334 CCTTTATGAT.....AACATATGATTAATTCCTTTTCTTCTTAT 6294
90 uLeuGlnLeuTyrLeuAlaThrTyrLeuGlyValPheTyrCysAlaLys 106
|||
6293 TTTTCCCTTAATTTTTCATTAATTTTTCCTTAATTTTTCCTTAATTA 6244
107 ValAlaSerValArgHisProLeuPheIleTyrLeuLysMetArgIleSe 123
|||
6243 .....CCATTTCTTCTATG..... 6229
123 rLysLeuValProTrp.....MetIleLeuGlySerLeuLeuTyrAla 138
|||
6228 .....CCCTTCATGATGATGCTCTCTTTTTCATTTGTTTACCAATA 6189
138 eRMetIleCysValPheHisSerLysTyrAlaGlyPheMetValProTyr 154
|||
6188 GCATCATCTCTCATTT.....GTACTTTCACGATTTG 6157
155 PheLeuArgLys.....PhePheSerGlnAs 163
|||
6156 TTTTAACGATTAATGTTGTATGATCATCATTAATTTTTCAT..... 6112
163 naIleThrIleGlnLysGlnsPheIleAlaIleGlnIlePheSerPheY 180
|||
6111 .....ATACATCATTTTTCATAAT 6093
180 aLalagLupheSerValProLeuLeuIlePheLeu..... 191
|||
6092 TATCATTTCTATCTTCATTTACAAAGCATACATCATTAATTAATAATA 6043
192 .....PheAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrAr 206
|||
6042 GAACAACTCATTAATTAATTAATTAATTAATTAATTAATTAATTTTCG... 5995
206 gGlnMetArgAsnThrValAlaGlySerArgValProGlyArgGlyAlaP 223

```

```

5995 ..... 5995
223 rolleSerAlaLeuLeuSerIleLeuSerPheLeuIleLeuTyrPheSer 239
|||||
5994 .....TTATCTATTAAGTCTTTCATCTTAATTAATTTGTAT 5959
240 HisCysMet.....IleLysValPheLe 247
|||
5958 TCATTAATTAATTAAGCATTAACAAATTAATTAATTAATTAATTAATTA 5909
247 uSerSerLeuLysPheHisIleArgArgPheIlePheLeuPhePheIle 264
|||
5908 ATTGATTAATTAATTTTTCCTTAATTAATTAATTAATTAATTAATTAAT 5859
264 euValIleGlyIle..... 268
|||
5858 TCCTATTTTCTCATCTGATGATTAATTAATTAATTAATTAATTAATTAAT 5809
269 .....TyrProSerGlyHisSerLeuIleLeuIleLeuG 280
|||
5808 TTACTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5759
280 yAsnProLysLeuLysGlnAsnAlaLysLysPheLeuLeuHisSer 295
|||
5758 TAATCCCATATCTTATTAATTAATTTTCTTAATTAATTAATTAATTAATTA 5713

seq_name: /cgn2.6/ptodata/2/lna/5A_COMB.seq:US-08-056-200-93
seq_documentation_block:
; Sequence 93, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinhert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichophyllin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9551 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO

```

```

; NAME/KEY: INCL01
; LOCATION: 1645..251
;

```



Quality: 104.50 Length: 349  
Ratio: 0.611 Gaps: 21  
Percent Similarity: 48.997 Percent Identity: 23.496

Alignment\_block:  
US-09-510-332-1 x US-08-137-175A-7/rev ..

Align seg 1/1 to reverse of: US-08-137-175A-7 from: 1 to: 1958

```

1 MetLeuGluSerHisLeuIleLeuPheLeuLeuAlaValIle..... 15
   |||||
1034 GTTTGGCTCAGCACCCTTTTGTCCACATGCTATTAAAGCTAATACCTAA 985
   |||||
16 GluPheLeuGluGlyIlePheThrAsnGlyIleIleValValAlaG 32
   |||||
984 GCAATTCCTAGTAATATTTTTC.....TAA 956
   |||||
32 LysIleAspLeuIleLysHis.....Arg 39
   |||||
955 CTCTCCCTAATTTTAAAGCATCTTAAAGTCTTAAAGCTTGAATTTTGA 906
   |||||
40 LysMetAlaProLeuAspLeuLeuSerCysLeuAlaValSerArgI 56
   |||||
905 CTGCTTGGCTCTAGATTGG.....TGCCTG..CTGAGTCGAT 868
   |||||
56 ePhe.....LeuGluLeuPheIlePheT 64
   |||||
867 TTTTGTACTGTTATGTGCTCTTTTGTGAATACAGCTTTTGGTTT 818
   |||||
64 yr.....ValAsnValIleVal 69
   |||||
817 TCGGCTATTACACTAATGTTAAAGTGAAGCTTTGAAATCCATGTC 768
   |||||
70 IlePhePheIleGluPheIleMetCysSerAlaAsnCysAlaIleLeu 86
   |||||
767 CAGTTTTTTATAGAGCTGAGTAGTGTAGAGCATTAAGCTCAACTGTT 718
   |||||
86 upheIleAsnGluLeuGluLeuPheLeuAlaThrTTPleuGlyValPheT 103
   |||||
717 ATTTCTCCAGAGTTGAAATGT.....G 695
   |||||
103 yrcysAlaLysValAlaSerValArgHisProLeuPheIleTTPleuLys 119
   |||||
694 TTTGCTTAAACAAACAGTCCCTCTGTAACCTTTAATGTTGTTGCCG 645
   |||||
120 MetArgIleSerLysLeuValProTTPMetIleLeu..GlySerIleLeu 135
   |||||
644 CAGCAGCTAGAGTTCCCTTCAAGAGCAAGCTTTTAAACTCTTTAGCT 595
   |||||
136 TyrValSerMetIleCysValPheHisSerLysTyrAlaIleGlyPheMet 152
   |||||
594 TTT..CCGGTTTATCGCTTTTATGCTGTGATTTCAAGT...CTGGT 551
   |||||
152 LProTyrPheLeuArgLysPhePheSerGlnAsnAla.....T 165
   |||||
550 TCCATTGCTCTTACTATGTTTTCAGATGCTTCACCTTGGCTTGA 501
   |||||
165 hrIleGlnLysGluAspThrLeuAlaIleGlnIlePheSerPheValAla 181
   |||||
500 ATTTTCTTCTGTTGATGACTTGTCTTAAAGGTTACTTTTGTGATACT 451
   |||||
182 GluPhe.....SerValProLeuLeu..... 188
   |||||
450 AATGTTTGGCATCTCTTGAAGATTCAAAATGTGGTTTACTTAATC 401
   |||||
189 .....IlePheLeuPheAlaValLeuLeuIlePheSerLeuGly 203
   |||||
400 CTCAGCATTTGTTAATTTGCTTACTTTGTCAAGTTTTC..... 358
   |||||
203 rghIstHrArgGlnMetArgAsnHrValAlaGlySerArgValProGly 219
   |||||
357 .....CCTTCAAGTGTTCAGAA 340

```

```

220 .....ArgGlyAlaProIleSerAlaLeuSerIleLeuSe 232
   |||||
339 CCGTGTGTTTATACAGAGTTCTTTAAGCTCAAGCTTGTCTACTGTCC 290
   |||||
232 rPheLeuIle..LeuTyrPheSerHisCysMetIleLysValPheLeuSer 248
   |||||
289 CATTAGACTATATTAC.....CATCTTGTCTTTTCTTACTTACAA 246
   |||||
249 SerLeuLysPheHis.....IleArgArgPheIlePhe..... 259
   |||||
245 GAACCTTCATTTCCACAGTAAATCTACTGAAACGCTATTTTTTCATCA 196
   |||||
260 .....LeuPhePheIleLeuValIleGlyIleT 269
   |||||
195 AGCGTGTACATTTCCTTACATGCTATTAAATGCTAATTTAGACTAT 146
   |||||
269 yrPro.....SerGlyHisSerLeuIleLeuIleLeu 279
   |||||
145 TCCCAATAAATATTTTTCATATATATTTCTCCTTTTATATATA 102
   |||||
seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:us-08-479-017-7
seq_documentation_block:
; Sequence 7, Application US/08479017
; Patent No. 6143872
; GENERAL INFORMATION:
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: BERGSTROM, Sven
; TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDOFFERI AND
; TITLE OF INVENTION: PROPHYLAXIS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEWMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,017
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/137,175
; FILING DATE: 26-OCT-1993
; APPLICATION NUMBER: PCT/US92/08972
; FILING DATE: 22-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: BARBOUR-1B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Borrelia burgdorferi
; STRAIN: Ip90
; INDIVIDUAL ISOLATE: Isolate from I. persulcatus from
; INDIVIDUAL ISOLATE: Soviet Union

```

2

SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs



```

: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US-08-487-826B-13

```

```

alignment_scores:
  Quality: 104.00      Length: 331
  Ratio: 0.588         Gaps: 15
  Percent Similarity: 53.474   Percent Identity: 21.148

```

```

alignment_block:
  US-09-510-332-1 x US-08-487-826B-13

```

```

Align seg 1/1 to: US-08-487-826B-13 from: 1 to: 18124

```

```

1 MetLeuGluSerHisLeuLeileIeTyRpheLeuLeuAlaValIleGlnPhe 17
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17446 AIGCTTAAATCATCTATCATGATGATATATGATGATCATATATATGTC 17495
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17   eLeuLeuGlyIlePheThAsnGlyIleIleValValValAsnGlyIleA 34
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17496 ACTAATAATAATATGATGATGATGATGATGATGATGATGATGATGAT 17545
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 sPheuLeuHisArgLysMetAlaPheLeuAspLeuLeuSerCys 50
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17546 TGTAAATAATAATAAAGAGATTTTTCAGAGAGATATCTCATATATGATA 17595
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 LeuAlaValSerArgLeu.....PheLeu.....GlnLeu P 61
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17596 TATGGAATATCTAAATATATATATATCTTTTGTGTGTCATATATATAT 17645
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 heilePheTyRValAsnValIleValIlePhePheIleGluPheIleMet 77
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17646 TTGTAATATATGATATGATGATGATGATGATGATGATGATGATGAT 17695
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
78 CysSer...AlaAsnCysAlaIleLeuLeuPheIle...AsnGluLeuG 92
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17696 TGTITTTGGTATATTTGTAAAAATATCTTTTGTATATATATATATAT 17745
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
92 uLeuTyRLeuAlaThTyRLeuGlyValPheTyRCysAlaLysValAlaAs 109
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17746 TATATTTTATATATTTGCAACATGATTTTCTTTCTTTCTTATTTGTC 17795
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 eValAlaGHisPheLeuPheIleTyRLeuLysMetArgLysSerLysLeu 125
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17796 TAAATTTTTCATATAATTTATATATATATATATATATATATATAT 17830
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126 ValProTyRMetIleLeuGlySerLeuLeuTyRValSerMetIle..... 140
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17831 .....TTTATTTTATATATATATATATATATATATATATATATAT 17874
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
141 CysValPheHisSerLysTyRAlaGlyPheMetValProTyRLeuLeu 157
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17875 TTATGTATATATATGTTATTTGTTGTTCTTTTCTTGTACATTTTTCGAA 17924
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
157 rGlySPhePheSerGlnAsnAlaThrIleGlnLysGluAspThrLeuAla 173
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17925 TATATATATATAT.....ATATATATATATATATATATATATATAT 17959
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
174 IleGlnIlePheSerPheValAlaGluPhe.....SerValProLeuLe 188
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17960 CTAAATATATATATATCTCTATTTTTCATTTTTCATTTTTCAGATAT 18009
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
188 uIlePheLeuPhe..... 192
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18010 AATTTATTTTATTTTATTTGATATTTTATATATATATATATATATATG 18059
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
193 .....AlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArg 206
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

18060 TTTATATATGCTTTTATATATGCTTTTATTTT.....GTTACT 18100
207 GlnMetArgAsnThrValAlaGlySerArgValProGlyArgGlyAlaP 223
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18101 CTATATCTGAATATATCCAGCGAAAAAAATATATATATATCTC..... 18142
223 oIleSerAlaLeuLeuSerIleLeuSerPheLeuIleLeuTyR..... 237
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18143 .ATATATAAATTTATTTATATATATATATATATATATATATATATAT 18191
238 .....PheSerHisCysMetIleLysValPheLeuSerSerLeu 250
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18192 AATTAATATATATATATATATATATATATATATATATATATATATAT 18241
251 LysPhe.HisIleArgArgPheIlePheLeuPhePheIleLeuValIleG 267
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18242 AATTTCTTATTTTATTTTAACTTATTCCTTTTAAATTTCTTATATCTT 18290
267 LylleTyRProSerGlyHisSerLeuLeuLeuLeuGlyAsnProLys 283
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18291 ..TATCAACAAAAACATAAAGTATATCTCATATATCAACAAAAAA 18338
284 LeuLysGlnAsnAlaLysLysPheLeuLeuHisSerLys 296
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
18339 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 18377
seq_name: /cgn2_6/ptodata/2/lna/6_COMB.seq:US-08-748-506-9

```

```

seq_documentation_block:
  Sequence 9, Application US/08748506
  Patent No. 6159707
  GENERAL INFORMATION:
  APPLICANT: Ronnett et al.
  TITLE OF INVENTION: NOVEL SPERM RECEPTORS
  NUMBER OF SEQUENCES: 31
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Leydig, Volt & Mayer, Ltd.
  STREET: Two Prudential Plaza, Suite 4900
  CITY: Chicago
  STATE: IL
  COUNTRY: US
  ZIP: 60601-6780
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/748,506
  FILING DATE: 08-NOV-1996
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 60/033,751
  FILING DATE: 09-NOV-1995
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
  REFERENCE/DOCKET NUMBER: 74940
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 312-616-5600
  TELEFAX: 312-616-5700
  INFORMATION FOR SEQ ID NO: 9:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 984 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  US-08-748-506-9

```

```

alignment_scores:
  Quality: 103.50      Length: 265
  Ratio: 0.690         Gaps: 12

```



420 ..... 420  
74 uphelleMetCysSerAlaAsnCysAlaIleLeuLeuPheIleAsnGlu 91  
421 .....GCTGACCTGACCTTCGTGGTGACGC 445  
91 euGIuLeuTriPheuAlaThrTrp.....Leu 99  
446 TCCCCCTGTGG...GCTACCTACAGCTACCGGACTAGACTGGCCCTTT 492  
100 GlyValPheTyrCys..... 104  
493 GGGACCTTCTTCTGCAGCTGACAGTACCTCATCTTGTCAACATGTA 542  
105 .....AlaLysValAlas 109  
543 CCGCAGCGTCTTCTGCTCACCGGCTCAGCTTGACCGCTACCTGGCCA 592  
109 exValArgHisProLeuPheIleTrPLeuLysMetArgIleSerLysLeu 125  
593 TCGTGAAGCCCATGGCCATGCTCGGCTGAGGCTGGGGTGACGGGGCC 642  
126 ValProTrpMetIleLeuGlySerLeuLeuTyrValSerMetIleCysVa 142  
643 GTGGCCACGCGAGTCTTGGGTGGCGCGCCCTCTGGCCATGCTGT 692  
142 lPheHisSerLysTyrAlaGlyPheMet.....ValProT 154  
693 CATGGTGTATACGACCAACCGGACTTGGAGAACACCACTAAGTGCAGT 742  
154 YrPheLeuArgLysPhePheSerGlnAsnAlaThrIleGlnLysGluAsp 170  
743 GCTACATGGAC.....TACTCATGTGGCGCCTGAGCTCAGAGTGG 786  
171 Thr.....LeuAlaIleGlnIlePheSerPheValAlaGluPheSerVa 185  
787 GCCTGGAGGTGGGCTTGGGTCTGCTCCACCAACCGTGGCTTGTGGT 836  
185 lProLeuLeuIlePheLeuPheAlaValLeuLeuLeuIlePheSerLeuG 202  
837 GCCCTTACCATCATGCTGACGCTGTACTTCTTCATGCCCCAACCATCG 886  
202 lYargHisThrArgGlnMetArgAsnThrValAlaGlySerArgValPro 218  
887 CTGGCCACTTCCGCAGAGAACGC.....ATCGAGGGCTCGGAGACGG 930  
219 GlYArgGlyAlaProIleSerAlaLeuLeuSerIleLeuSerPheLeuI 235  
931 CGCCGG.....CTGCTCAGCATCATCGTGGTGTGGT 962  
235 eleuTyrPheSerHisCysMetIleLysValPheLeuSerSerLeuLysP 252  
963 GGTGACCTTTCCTGCTGCTGG.....ATGCCCT 991  
252 heHisIleArgArgPheIlePheLeu..... 260  
992 ACCACCTGTGTAAGACGCTGTACATGTGGGACGACCTGCTCACTGGCCC 1041  
261 .....PhePheIleLeuValIleGlyIleTyrPro..... 270  
1042 TGTGACTTTGACCTTCTTCTCATGAACTTCCCTACTGCACTGCAT 1091  
271 .....SerGlnHisSerLeuIleLeuIleLeuGlyAsnP 282  
1092 CAGTACGTCAACAGCTGCTCAACCCCTTCTATGCTTTTTCGAC 1141  
282 polyLysLeuLysGlnAsnAlaLysLysPheLeuLeuHisSerLysCysCys 298  
1142 CCGGCTTCGCGCAGGCTGCACTGCATGCTC.....TGTGT 1179

**THIS PAGE BLANK (USPTO)**







[illegible]

736 .....GTTCACTTTTCTACTTTCTCTAAACT 708

[illegible]

**343** Jolyvetal phorou      CoeCoef eurwobhavet1 e-      • unob      oes

65 / TTTTCAACATATTTTTCATCTTGTCTCTT 608

```
z08 lemeleupepneileualeilleillyproserglyhis 274      |||
:::|||||||:::   ::|
607 CCCCCCCCCCCCCCCCTCCATCAGGACCATTCATTTTCCCCTGGCGA  |||
```

[illegible]

6/2 negative negative 4/2

205 AC11C11G11C11 348

seq\_name: /cgnz\_2/gcdata/geneseq/NA1996.DAT:T17115

seq\_documentation\_block:  
ID T17115 standard: DNA: 3016 BP

AA  
AC  
T17115:

06-MTH-1996 (£4284 000.00)

Phodopsin gene  
XX  
DE

Human: rhodococcus ruber

KW probe; primer; hybridisation; polymerase chain reaction; po

H  
C  
E  
C  
C  
C  
C  
C  
C

X  
X  
C  
C

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

ET	5'0TR	200.294
ET		

Feature	5' UTR	202..294
FT		
EE		

```

FT /note= "Alternative 5'-UTR"

```

```
ET      /*tag= c
```

FT	CDS	295..5278
----	-----	-----------

```
/product= Rhodopsin
```

FT / \*tag- e

FT (normal) "

```
FT      /*tag= f
```

[illegible]

ET /\*tag= g

```
FT primer_bind 362..381
```

333





```

XX  BacRI: bacteriocin; antimicrobial; antibacterial;
KM  Moraxella bovis: Infectious bovine keratoconjunctivitis; cancer;
XX  therapy; ds.
XX
OS  Staphylococcus aureus strain U70007 (ATCC 55800).
XX
PN  WO9812319-A1.
XX
PD  26-MAR-1998.
XX
PF  18-SEP-1997; 97WO-US16758.
XX
PR  17-SEP-1997; 97US-0931999.
XX  19-SEP-1996; 96US-0710561.
XX
PA  (UNIV ) UNIV KANSAS STATE RES FOUND.
XX
PI  Crupper SS, Iandolo JJ;
XX  WPI: 1998-230316/20.
XX
DR  Therapeutic proteinaceous substances from Staphylococcus aureus -
XX  useful to inhibit growth of wide range of prokaryotic or eukaryotic
XX  cells, e.g. Moraxella bovis causing infectious bovine
XX  keratoconjunctivitis
XX
PS  Claim 2; Page 19-23; 38pp; English.
XX
CC  This polynucleotide comprises the bacteriocin BacRI operon of
CC  Staphylococcus aureus U70007. The sequence of the BacRI operon
CC  was determined by N-terminal sequencing of purified BacRI peptide
CC  (see W54171), with back-translation and plasmid analysis. The
CC  BacRI operon includes the BacRI gene (see V21510), a homologue of
CC  the cym gene of the cytolysin operon of Enterococcus faecalis
CC  whose function is involved in the maturation of pre-cytolysin,
CC  an ATP-transporter gene, b1ol and b1oz genes related to
CC  lactococin biosynthesis and modification, and a gene involved in
CC  immunity function. BacRI peptides can be produced by construction
CC  of an expression vector containing an oligonucleotide or operon
CC  coding for BacRI, and use of the vector to transform host cells for
CC  BacRI expression. The entire BacRI operon has been cloned into
CC  plasmid pUB110, and Bacillus subtilis transformants secreted the
CC  recombinant BacRI peptide into the medium. Bacteriocin BacRI is
CC  active against many Gram-positive and Gram-negative organisms such
CC  as Bordetella bronchiseptica, Pasteurella multocida and
CC  Staphylococcus aureus; Moraxella bovis, causing infectious bovine
CC  keratoconjunctivitis, is especially sensitive. BacRI can also be
CC  used as an anti-cancer agent.
XX
S0  Sequence 6755 BP; 3903 A; 1252 C; 1600 G; 0 U; 0 other;

Alignment_scores:
      Quality: 124.50      Length: 238
      Ratio: 1.029      Gaps: 13
      Percent Similarity: 50.840      Percent Identity: 26.471

Alignment_block:
US-09-510-332-1 x V21511/rev ..

Align seg 1/1 to reverse of: V21511 from: 1 to: 6755

      48  leuSerCysLeuAlaValSerArgIlePheLeuGlnLeuPhe.....II 62
      |||||||:::|||||:::|||||
908  CTCCTCTGCGTGCCTTTTTCGTGCTTGGCTTTTTCGTGGCGCTT 859
      62  ePheTyrValAsnValIleValIlePhePheIleGlnPheIleMetCys 79
      ::|||:::|||||:::|||||
858  CTTTGTCTTTGGGGGTGTTTGTCTTTCGCTCTTCCTTCCTGTT 809
      79  eAlaAsnCysAlaIleLeuLeu.....PheIleAsnGlnLeuGlnLeu 93
      :::::|||||:::|||||:::

```

[illegible]





```

KM hereditary; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FH prim_transcript 200..1341
FT /*tag= a
FT CDS 295..1341
FT /*tag= b
XX
XX MO9312134-A.
XX PD 24-JUN-1993.
XX PF 08-DEC-1992; 92MO-US10536.
XX PR 11-DEC-1991; 91US-0805123.
XX PA (HARD ) HARVARD COLLEGE.
XX PI Berson EL, Dryja TP;
XX WPI: 1993-214088/26.
XX DR P-PSDB; R38483.
XX
XX Probe or primer conty. sequence of human retinal degeneration
XX slow protein mutant - used to diagnose hereditary retinal
XX degenerative diseases
XX
XX Disclosure: fig 1; 56pp; English.
XX
XX The sequence given represents the human rhodopsin cDNA. Mutant
XX versions of this sequence encode proteins which cause retinal
XX degeneration. These sequences may be identified using primers/
XX probes described in the invention (see also Q43545-48) and may be
XX used to diagnose hereditary retinal degeneration. This sequence is
XX CC the closest approximation to the gene sequence as the sequence given
XX in the specification is not printed clearly.
XX
XX Sequence 3129 BP; 694 A; 999 C; 654 G; 735 T; 47 other;
XX
XX
XX alignment_scores:
XX Quality: 122.50 Length: 327
XX Ratio: 0.756 Gaps: 16
XX Percent Similarity: 49.541 Percent Identity: 21.101
XX
XX alignment_block:
XX US-09-510-332-1 x Q43543 ..
XX
XX Align seg 1/1 to: Q43543 from: 1 to: 3129
XX
XX 1 MetLeuGIuSerHISLeuIleIleTyPheLeuLeuAlaValIleGlnP 17
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 409 ATGCTGCGCCGCTACATG.....TTTCTGCTGATCGTCGCGGCTT 449
XX
XX 17 eLeuLeuGIlyIlePheThrAsnGlyIleIleValAlaValAsnGlyIleA 34
XX | :|||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 450 CCCCATCAACTCTCTACGCTCTACGCTACGCTC..... 483
XX
XX 34 splLeuIleLysHisArgLysMet...AlaProLeuAspLeuLeuSer 49
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 484 .....CAGCACAAGAGCTGCGCACCCTCTCACTCACTCACTCTGTC 525
XX
XX 50 CysLeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyValAs 66
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 526 AACCTAGCGCTGCTGACCTCTTCATGCTGCTGCTGCTGCTGCTGCTG 575
XX
XX 66 nValIleValIlePhePheIleGluPheIleMetCysSerAlaAsnCysA 83
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 576 CCTTACACCTCTCTGATGATGATGATGATGATGATGATGATGATGATG 625
XX
XX 83 lAlleuLeuPheIle.....AsnGluLeuGlnLeuTrp... 94

```

```

626 ATTTGARGGCTTTTCCACCCTGGCGGTGAATATGCCCCCTGCTGC 675
XX
XX 95 .....LeuAlaThrTrpLeuGlyValPheTyCysAlaLysValAl 108
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 676 TTGGTGCTCTGGCCATCGAGCGGTACGTGTGTGTGTGTGTGTGTGTG 725
XX
XX 108 aserValArg.....HisProLeu.....PheIleI 117
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 726 CAACCTCCGCTTCGGGAGAGAACCATCATCATGCGCTTCCTTCACT 775
XX
XX 117 TP.....LeuLysMetArgLysSerLysLeuValProTrpMetIle 130
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 776 GGTTCATGCGCGCTGCTGCGCGCCACCCCACTCGCGCGCTG..... 819
XX
XX 131 LeuGlySerLeuTyValSerMetIleCysValPheHisSerLysTy 147
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 820 .....TCACAGTA 827
XX
XX 147 rAlaGlyPheMetValProTyPheLeuArg.....LysP 159
XX | :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 828 C.....ATCCCCAGGCGCTGCACTGCTCGTGGATTCGACT 865
XX
XX 159 hepSerGlnAsnAlaThrIleGlnLysGlnAspThrLeuAlaIleGln 175
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 866 ACTTACACCTCAAGCCGAGGTCAACACGAG.....TCCTTGTGTC 906
XX
XX 176 IlePheSerPheValAlaGluPheSerValProLeuLeuIlePheLeuP 192
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 907 ATCTACATGTTCTGTGTCACCTCCATCCCATGATATATCATCTTTT 956
XX
XX 192 eAlaValLeuLeuLeuIlePheSerLeuGlyArgHisThrArgLysMet 209
XX | :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 957 CTGCTATGCGCAGCTGCTGCTTACCGTCAAGAGCGCGCTGCCAAGC 1006
XX
XX 209 rGAsnThrValAlaGlySerArgValProGlyArgGlyAlaProIleSer 225
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1007 AGGAGTCA...GCCACACACAGAGCAGAGAGAGAGAGAGAGAGAGAG 1053
XX
XX 226 AlaLeuLeuSerIleLeuSerPheLeuIleLeuTyPheSerHisCysMe 242
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1054 GTCATCATCATGTCATGCTTCTGATCTGCTGCTGCTGCTGCTGCTGCTG 1103
XX
XX 242 tIleLysValPheLeuSerSerLeuLysPheHisIleArgArgPheIleP 259
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1104 CGTGCATCTTACATCTTCAACCCAGCGCTCAACTTGGTCCCACT 1153
XX
XX 259 heLeu.....PhePheIleLeuValIleGlyIleTyTrpSer 271
XX |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1154 TCATGACCATCCAGCGCTTCTTGGCCAGTCCGCGCCCATCTCAACCT 1203
XX
XX 272 GlyHisSerLeuLeuLeuLeuGlyAsnProLysLeuLysGlnAsnAl 288
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1204 GTCATCATATCATATGATATC.....AACCTTCCGGAACATG 1241
XX
XX 288 aLysLysPheLeuLeuHisSerLysCysCys 298
XX :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1242 C.....ATGCTCACCCATCTGCTGC 1263
XX
XX seq_name: /cgn2.2/gcdata/geneseq/geneseq/NM1998.DAT:V31718
XX
XX seq_documentation_block:
XX ID V31718 standard; DNA; 1202 BP.
XX
XX AC V31718;
XX
XX 11-SEP-1998 (first entry)
XX
XX DE Kaposi's sarcoma associated herpesvirus ORF74.
XX
XX PCR primer: KSHV; ORF74; Kaposi's sarcoma; diagnosis; treatment;
XX
XX G protein-coupled receptor; ss.
XX
XX

```



FT	CDS	complement (69412..69915)
FT		/*tag= h
FT		/product= glycoprotein L
FT	CDS	complement (88410..88910)
FT		/*tag= 1
FT		/product= interferon regulatory factor 2
FT	CDS	89600..90341

```

FT CDS 90173..90643
FT /*tag= k
FT /product= glycoprotein X
FT /complement (93636..94127)
FT /*tag= 1
FT /product= interferon regulatory factor 4
FT /complement (11931..112443)
FT /*tag= m
FT /product= casp1 protein IV
FT /complement (123808..127296)
FT /*tag= n
FT /product= immediate early protein

MO9804576-A1.
PD 05-FEB-1998.
XX 22-JUL-1997; 97WO-US13346.
XX 29-NOV-1996; 96US-0757669.
XX 25-JUL-1996; 96US-0686243.
XX 25-JUL-1996; 96US-0686349.
XX 25-JUL-1996; 96US-0686350.
XX 25-JUL-1996; 96US-0687253.
XX 25-JUL-1996; 96US-0688814.
XX 05-SEP-1996; 96US-0708678.
XX 10-OCT-1996; 96US-0728323.
XX 13-NOV-1996; 96US-0747887.
XX 13-NOV-1996; 96US-0748640.
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
XX WPI: 1998-130615/12.
XX
XX New nucleic acid encoding Kaposi's sarcoma associated herpes virus
XX proteins - useful for e.g. detecting levels of HHV8 in, and
XX preparation of vaccines for treatment of, HIV patients
XX
XX Example 2; Page 135-203; 230pp; English.
XX
XX This sequence represents the long unique region and terminal repeat of
XX the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
XX as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
XX invention which encode KSHV polypeptides selected from: (a) viral
XX microphage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
XX (c) viral INF 1; (d) complement-binding protein; glycoproteins B, M or L;
XX (d) capsid protein IV encoded by ORF5; and (e) immediate early protein
XX encoded by IL, and antibodies (Ab) specific for the nucleic acid, proteins encoded
XX by IL, and antibodies (Ab) specific for the proteins are useful for
XX detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
XX fluids or tissue samples. HHV8 infections can be treated with antisense
XX or triplex forming molecules or agents that bind specifically to the
XX protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
XX to differentiate between lymphomas, and HHV8 may be implicated in many
XX other lymphoproliferative diseases such as lymphomas, leukaemia,
XX splenomegaly and mycosis fungoides. Cells and animals containing the
XX nucleic acid are useful for drug screening. HHV8-derived peptides can be
XX used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
XX can be inhibited with methotrexate. These can also be used to determine
XX the immune status of a patient infected with HIV. HHV8 derived protein
XX viral MIP II may be used as an anti-inflammatory agent for
XX e.g. treating rheumatoid arthritis. This sequence is stated as containing
XX 81 open reading frames.
XX
XX Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

```

Alignment\_scores: Quality: 120.50 Length: 263  
Ratio: 0.867 Gaps: 14

```

Percent similarity: 52.852 Percent identity: 22.433
Alignment block:
US-09-510-332-1 x V19941 ..
Align seg 1/1 to: V19941 from: 1 to: 137507

11 LeuLeuAlaValAlIeGlnPheLeuLeuGlyIlePheRhrAnGlyIleI 27
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
129527 ATACTCTCTGCAAT...TTCTCATTAATGTTCTTGAATGATTCGT 129573
27 eValValAlaValnGlyIleAspLeuIleLysHisArgLysMetAla...P 43
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129574 CACCTACAT...TTTGCACACCGATCCGCGGACGAG 129611
43 rOlAuAspLeuLeuLeuSerCysLeuAlaValSerArgIlePheLeuGln 59
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129612 CGATAGATATATCTCTCTCTGATCTGCTTAACTGCTGCTTAAAGC 129661
60 LeuPheIlePheTyValAlaValAlIeValIlePhePheIleGluPheI 76
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129662 ATATCTCTA...TTGCCAGAGTGTGATGATGTTTGTTCACATAT 129705
76 eMeCySerAlaAsnCysAlaIleLeuLeuPheIleAsnGluLeuGlu 93
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129706 CATCTCCACAGGCTTGTGACAGACTTGAAATTTT...T 129743
93 eUrTrPLeuAlaTrHrTrPLeuGlyValAlPheTyrcysAlaValAlaSer 109
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129744 ACTATTATATGTCTACTGATATCTTCAAGTGTGTCGTCAGATCTA 129793
110 ValArgHisProLeuPheIle...TrPLeuLysMetAr 121
|||||:::|||||:::|||||:::|||||:::|||||:::
129794 GTGAGTACTCTCTCTGCGGATTTCTACCGCTTCTGCGCCAG... 129838
121 gIleSerLysLeuValProTrPMeIleLeuGlySerLeuLeuTyValS 138
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129839 ...AMCAGCTCCCTCGATGAGTACTGATGATCCGCTGACCTTAAATG 129884
138 eMeIle...CysValPheHisSerLysTyAlaGly 149
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129885 CATTTGGCTGTGCGGGGATCCCTGTGCACAGCAGCGGTGTCGAC 129934
150 PheMetValProTyrcPheLeuArgLysPhePheSerGlnAsnAlaThrI 166
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
129935 ...CCGTCACGACAGCGCCATCTTTAGACACCGCGGAA 129975
166 eGlnLysGluAsp...ThrLeuAlaIleGlnIlePheSerPheValAlaG 182
:::|||||:::|||||:::|||||:::|||||:::|||||:::
129976 CATGACTGCAGACTGCGACATGTCAGAACCGTGCATGATTCGACG 130025
182 lUpHeserValProLeuLeuIlePheLeuPheAlaValLeuLeuLeu 198
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
130026 GTTCTCTGTACCCCTG...GCCCTCTTATTCGTTT 130060
199 PheSerLeuGly...ArgHisThrArg...GlnMetArgAs 210
:::|||||:::|||||:::|||||:::|||||:::|||||:::
130061 TATGCTCTCACTGTGTGTGTGAGAGAGACAAAGCTGACCCAGCGG 130110
210 nTrValAlaGlySerArgValProGlyArgGlyAlaProIleSerAla 227
:::|||||:::|||||:::|||||:::|||||:::|||||:::
130111 GAAGGTAAGGGG... 130123
227 eUleuSerIleLeuSerPheLeuIleLeuTyrcPheSerHisCysMetIle 243
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
130124 ...GTGATGTGTGCTGTGCTGTGCTGTTTGTGTGTTTGTCTTCCT 130168
244 LysValPheLeuSerSerLeuLysPheHisIleArgArg 256
|||||:::|||||:::|||||:::|||||:::|||||:::
130169 TACACGTAATTAATCTACTGACACTGTGTAAGCGCA 130207
seq_name: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT.T79064

```





```

ID T79066 standard; cDNA to mRNA; 1147 BP.
XX
XX T79066;
XX
XX 10-NOV-1997 (first entry)
XX
XX Xenopus melatonin receptor MEL-1A short form coding sequence.
DE
XX
XX Allele: Xenopus laevis; melatonin; receptor; untranslated region; PCR:
XX mRNA; half-life; skin; amplification; primer; polymerase chain reaction;
XX transmembrane domain; cellular signaling; inhibition; adenylyl cyclase;
XX modulation; intracellular; cyclic GMP; inhibitor; phosphodiesterase; ss.
XX
XX Xenopus laevis.
OS
XX
XX Key Location/Qualifiers
XX FH 1.1065
XX FT CDS
XX FT /*tag= a
XX FT /product= melatonin receptor MEL-1A
XX
XX W09704094-A1.
XX
XX 06-FEB-1997.
XX
XX 24-JUL-1996; 96MO-FR01167.
XX
XX 24-JUL-1995; 95FR-0008947.
XX
XX (ADIR.) ADIR & CIE.
XX
XX Jockers R, Marullo S, Strosberg AD;
XX
XX WPI; 1997-132635/12.
XX
XX P-PSDB; W25927.
XX
XX New nucleic acid encoding functional melatonin receptor of Xenopus -
XX PT for screening for potential (ant)agonists useful for e.g. treating
XX PT cardiovascular disease and cancer
XX
XX Claim 1; Page 33-34; 62pp: French.
XX
XX
XX Sequences T79063-66 represent novel allelic genes of the Xenopus laevis
XX melatonin receptor MEL-1A. The sequences encode proteins which are 65
XX amino acids shorter than those described in the prior art. This sequence
XX is a short form of the novel receptor MEL-1A also known as
XX Mel-1-c(beta). As compared to the short form (T79066), the difference
XX occurs in the 3' untranslated region (both sequences encode the same
XX protein), which is thought to affect the half-life of the mRNA. The
XX MEL-1A sequences also differ from known MEL-1A receptor sequences by
XX 6 amino acids. The nucleotide sequence was isolated from cDNA derived
XX from Xenopus skin RNA and amplified using the primers T79067-76. The
XX nucleotide sequence encodes a protein which is a 7 transmembrane receptor
XX involved in cellular signaling. MEL-1A has been shown to modulate
XX intracellular cGMP, esp. inhibiting its accumulation induced by an
XX inhibitor of phosphodiesterase, but unlike MEL-1Aa (T79063) cannot
XX inhibit adenylyl cyclase activity.
XX
XX Sequence 1147 BP; 313 A; 246 C; 240 G; 348 T; 0 other;
XX
XX
XX alignment_scores:
XX Quality: 118.00 Length: 294
XX Ratio: 0.747 Gaps: 12
XX Percent Similarity: 53.741 Percent Identity: 19.388
XX
XX alignment_block:
XX US-09-510-332-1 x T79066 ..
XX
XX Align seg 1/1 to: T79066 from: 1 to: 1147
XX
XX 2 leuGluserHisLeuIleIleTyRpheLeuLeuAlaValIleGlnPheLe 18
XX ||| ||| ||| .....:
XX 91 CTCACCTCTGCGCCGTGGCGGTCTTATATTCACCACTGTGTGTGATGT 140
XX , , , , ,

```

```

18 uLeuGlyIlePheThrsnGlyIleIleValIleValAlaValAlaValIleAspL 35
XX ||||| .....:
141 CCTGGGC.....AATATATTGGTCATT.....TTGCTCG 169
XX
35 euIleLysHisArgLysMetAlaProLeu...AspLeuLeuLeuSerCys 50
XX :|||:.....:
170 TCCGTGAGGAAACAGACAGCTGAGAAATGCTGAAATCTCTTGTGTGTCAGT 219
XX
51 LeuAlaValSerArgIlePheLeuGlnLeuPheIlePheTyValaVala 67
XX |||:.....:
220 TTGCTATTGGCGATTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 269
XX
67 ILeValIlePhePheIleGluPheIleMeCysSerAlaAsnCysAlaI 84
XX :|||:.....:
270 CATAGCTATTTCAGAAATGGGTGACGCTTGGAATATTCATTCATTCAG 319
XX
84 ILeuLeuPheIleAsnGluLeuGluLeuTyRpheAlaThrTrpLeuGly 100
XX |||:.....:
320 TCAGTGGCTTCGTGATGGAGCTCAGCCTT.....ATTGGA 354
XX
101 ValPheTyRcysAlaLysValAlaSerValArgHisProLeuPheIleTr 117
XX :|||:.....:
355 TCAGTCTTCAACATACAGCCATGCTATGACACAGCTATTGCTACATCTG 404
XX
117 PLeuLysMetArgLysSerLysLeuVal.....ProTyrPheL 130
XX :|||:.....:
405 CCACAGCCTGAGATATGACAAAGCTTTTAAATCAAAAGACAGCCTGTTCT 454
XX
130 ILeuGlySerLeuLeuTyRValSerMetIleCysValPheHisSerLys 146
XX ||||| .....:
455 ACCTTGGCCTGACATGATGATGATGATGATGATGATGATGATGATGATG 504
XX
147 TyRAlaGlyPheMetVal.....ProTyrPheLeuArgLysPhePhe 161
XX :|||:.....:
505 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 554
XX
161 rGlnAsnAlaThrIleGlnLysGluAspThrLeuAlaIleGlnIlePhe 178
XX |||:.....:
555 GCAG.....ACGTAAGTTCCTCATGACCATACATACATGATGATGATG 598
XX
178 erPheValAlaGluPheSerValProLeuLeuIlePheLeuPheAlaVal 194
XX ||||| .....:
599 ATTTATATAGTCCCTTAAAGTGTGTGATGATGATGATGATGATGATG 645
XX
195 LeuLeuLeuIlePheSerLeuGlyArgHisThrArgGln..... 207
XX :|||:.....:
646 TGGGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 695
XX
208 .....MetArgAsnThrValAla..... 213
XX
696 AAAGTTCACACCAACAGACTTGAGAAATTTCTTGACCATGTTGTGTGTC 745
XX
214 .....GlySerArg 216
XX
746 TTGTACTTTTGGCGTTTGTGCTGGGACCCCTTGAAATTTTATGGCCTTGC 795
XX
217 Val.....ProGlyArgGlyAlaPro...IleSerAlaLeuLeuSe 229
XX ||| ||| ||| ||| |||
796 GTGGCCATTAAACCACTCCACGTGGCCACCAAGATTCAGAGATGGTGTGT 845
XX
229 rIleLeuSerPheLeuIleLeuTyRpheSerHisCysMetIleLysValP 246
XX :|||:.....:
846 TGTGTAAAGCTATTTCATGGCCTTATTTTAAACAGCTGTCTCAATGCTGT 895
XX
246 heLeuSerSerLeuLysPheHisIleArgArg 256
XX :|||:.....:
896 TCTACGCTGTGCTAAATCAAACTTCCGCAAG 927
XX
seq_name: /cgn2_2/gcgdata/geneseq/geneseq/NA1997.DAT:T79063
seq_documentation_block:
ID T79063 standard; cDNA to mRNA; 1311 BP.

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 11:39:04 ; Search time 65.02 Seconds  
(without alignments)  
82.577 Million cell updates/sec

Title: US-09-510-332-1

Perfect score: 1521  
Sequence: 1 MLESHTLIYFLAVIQFLG.....GNPKLNKAKKFLHSKCCQ 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/PCRTUS\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123.5	8.1	309	1	US-08-118-270-56
2	123.5	8.1	309	4	PCT-US93-08528-56
3	123.5	8.1	348	2	US-08-390-000A-8
4	122.5	8.1	401	2	US-08-820-521-2
5	120.5	7.9	327	3	US-08-748-506-23
6	120.5	7.9	327	3	US-08-748-506-23
7	120.5	7.9	342	3	US-08-785-928-1
8	120.5	7.9	342	3	US-08-728-603-17
9	118.5	7.8	524	2	US-08-928-692-12
10	118.5	7.8	420	2	US-08-466-103A-2
11	117.5	7.7	327	3	US-08-748-506-24
12	117.5	7.7	358	2	US-08-458-970A-11
13	107	7.0	286	4	US-08-118-270-65
14	107	7.0	286	4	PCT-US93-08528-65
15	104.5	6.9	359	4	US-08-148-209A-4
16	103.5	6.8	327	3	US-08-748-506-14
17	103	6.8	342	3	US-08-988-876-9
18	103	6.8	380	3	US-08-153-848-40
19	103	6.8	380	3	US-09-299-843A-40
20	103	6.8	380	4	PCT-US93-11153-40
21	100.5	6.6	316	4	US-08-118-270-44
22	100.5	6.6	316	4	PCT-US93-08528-44
23	100	6.6	277	4	US-08-118-270-68
24	99.5	6.5	277	4	PCT-US93-08528-68
25	99.5	6.5	359	2	US-08-748-485-5
26	99	6.5	325	1	US-08-118-270-51
27	99	6.5	325	4	PCT-US93-08528-51
28	99	6.5	395	4	US-08-097-938-5

29	99	6.5	395	1	US-08-476-000-5	Sequence 5, Appl1
30	99	6.5	395	1	US-08-472-840-5	Sequence 5, Appl1
31	99	6.5	395	2	US-08-476-976-5	Sequence 5, Appl1
32	99	6.5	395	3	US-08-474-410-5	Sequence 5, Appl1
33	99	6.5	399	1	US-08-476-000-61	Sequence 61, Appl1
34	99	6.5	399	1	US-08-472-840-61	Sequence 61, Appl1
35	99	6.5	399	2	US-08-476-976-61	Sequence 61, Appl1
36	99	6.5	399	2	US-08-474-410-61	Sequence 61, Appl1
37	98.5	6.5	372	1	US-08-202-056-5	Sequence 5, Appl1
38	98.5	6.5	372	1	US-08-076-093A-6	Sequence 5, Appl1
39	98.5	6.5	372	1	US-08-701-265-6	Sequence 6, Appl1
40	98.5	6.5	372	2	US-08-284-586-6	Sequence 6, Appl1
41	98.5	6.5	372	2	US-08-805-478-6	Sequence 6, Appl1
42	98.5	6.5	372	2	US-08-802-627A-6	Sequence 6, Appl1
43	98.5	6.5	372	2	US-08-801-238-6	Sequence 6, Appl1
44	98.5	6.5	372	2	US-08-801-228-6	Sequence 6, Appl1
45	98.5	6.5	372	3	US-09-104-296-6	Sequence 6, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-118-270-56  
Sequence 56, Application US/08118270  
Patent No. 5508384  
GENERAL INFORMATION:  
APPLICANT: Murphy, Randall B.  
APPLICANT: Schuster, David I.  
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
NUMBER OF SEQUENCES: 348  
CORRESPONDENCE ADDRESS:  
ADDRESS: BROWDY AND NEWMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/118,270  
FILING DATE: 09-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/943,236  
FILING DATE: 10-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, Kevin G.  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: MURPHY-2A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 309 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-118-270-56  
Query Match 8.1% Score 123.5, DB 1, Length 309;  
Best Local Similarity 21.1% Pred. No. 4.3e-05;  
Matches 69, Conservative 64, Mismatches 121, Indels 73, Gaps 17;  
OY 1 MLESHTLIYFLAVIQFLGIFTNGIIVVNGIDILIRKRM-APLDLLLSCLAVSRIFLQ 59

```

Db 2 MLAAYM---FLIIVGFPINFLTLVTV-----QHKRLRPIVYIILNLAVADLFMV 50
QY 60 LFIYVNVYIIFIEFIMCSANCAILLFINEL-----ELMLATWLGVPYCAKVASVR- 111
Db 51 LGGFTSTLYTLHGVEFVPGTCNLGFFATLGGELALMSLMLAIERYVYVCKPMSNFRF 110
QY 112 ---HPL---FIWLMKRISKLV-----WMILGSLVYVMICVFSKAGFVNPYFLR- 157
Db 111 GENHAIMGVAFVW-MALCAAPPIAGW-----SRY-----IPEGLOQSC 149
QY 158 --KEFSQNTIQKEDTLAIOISFVAEFSVPLILFPAVLLIFSLGRTRMRNTVAGS 215
Db 150 GIDYVTLKPEVNE---SPVIYMFVHFITPLILF-FCYQGLVFTVKEAAAOQES-ATT 204
QY 216 RVPGRGAPISALSLISFLIFSHCMIKVFLSLKPHIRFIPL-----FPIIVIGIYPS 271
Db 205 QKAKEVTRMVIIVIAFLICWVPYASVAFITFHQGSNFGPIFMRIAPFAKSAIY-- 262
QY 272 GHSLLILGNPKLONAKKFLHSKCC 298
Db 263 -NPVIYIIFNKQF---RNCMLQLIC 284

```

## RESULT 2

```

PCT-US93-08528-56
; Sequence 56, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BROWDY AND NETMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08528-56

```

```

Query Match 8.1%; Score 123.5; DB 4; Length 309;
Best Local Similarity 21.1%; Pred. No. 4.3e-05;
Matches 69; Conservative 64; Mismatches 121; Indels 73; Gaps 17;
QY 1 MLESHTLIYFLAVIOFLGIFNGIIVVNGIDLKHKRM-APDLLLSCLAVSRIFLQ 59

```

```

Db 2 MLAAYM---FLIIVGFPINFLTLVTV-----QHKRLRPIVYIILNLAVADLFMV 50
QY 60 LFIYVNVYIIFIEFIMCSANCAILLFINEL-----ELMLATWLGVPYCAKVASVR- 111
Db 51 LGGFTSTLYTLHGVEFVPGTCNLGFFATLGGELALMSLMLAIERYVYVCKPMSNFRF 110
QY 112 ---HPL---FIWLMKRISKLV-----WMILGSLVYVMICVFSKAGFVNPYFLR- 157
Db 111 GENHAIMGVAFVW-MALCAAPPIAGW-----SRY-----IPEGLOQSC 149
QY 158 --KEFSQNTIQKEDTLAIOISFVAEFSVPLILFPAVLLIFSLGRTRMRNTVAGS 215
Db 150 GIDYVTLKPEVNE---SPVIYMFVHFITPLILF-FCYQGLVFTVKEAAAOQES-ATT 204
QY 216 RVPGRGAPISALSLISFLIFSHCMIKVFLSLKPHIRFIPL-----FPIIVIGIYPS 271
Db 205 QKAKEVTRMVIIVIAFLICWVPYASVAFITFHQGSNFGPIFMRIAPFAKSAIY-- 262
QY 272 GHSLLILGNPKLONAKKFLHSKCC 298
Db 263 -NPVIYIIFNKQF---RNCMLQLIC 284

```

## RESULT 3

```

US-08-390-000A-8
; Sequence 8, Application US/08390000A
; Patent No. 5985583
; GENERAL INFORMATION:
; APPLICANT: Sealfon, Stuart C.
; TITLE OF INVENTION: Cloning and Expression of
; TITLE OF INVENTION: Gonadotropin-Releasing Hormone Receptor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,000A
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-052
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-390-000A-8

```

```

Query Match 8.1%; Score 123.5; DB 2; Length 348;
Best Local Similarity 20.8%; Pred. No. 4.9e-05;
Matches 68; Conservative 63; Mismatches 125; Indels 71; Gaps 16;
QY 1 MLESHTLIYFLAVIOFLGIFNGIIVVNGIDLKHKRM-APDLLLSCLAVSRIFLQ 59
Db 39 MLAAYM---FLIIVGFPINFLTLVTV-----QHKRLRPIVYIILNLAVADLFMV 87

```

QY 60 LFFEVNVIPIEFIEIMSCANCAILLFI-----NELEW-----LATWLGVCYCAKVASVR 111  
 DB 88 LGFSTLTYSLSHGIVFPGTCNLEGFATLGGELALMSLVLAERIVVYCKRPSNSR 147  
 QY 112 ----HPL----FTW----LKMRLSKLVPMMILGSLVYSMICVFSKYAGPMVPLR--- 157  
 DB 148 FGENHAIMGVAFWVWALACAPPLAGW-----SRY-----IPEGLOCSG 187  
 QY 158 --KFSQONATIQEDLAIQIESFVAEFSPPLIFLFAVLLIFESLGRHTRORNTVAGS 215  
 DB 188 GIDVYTLKPEVNE--SFVIYMFVVFHTIPMLITFCYGOVLFTYKAAQOQES-ATT 243  
 QY 216 RYDGRGAPISALSLISFLILYFSHCMIKIVLSLKFHRRIFL---FFILVIGYPS 271  
 DB 244 QKAKEKTRAVIIMVIAFLCWPYASVAFYIFTHOGSNGPIFMITPAFAKSAATY-- 301  
 QY 272 GHSLLILGNPKLKONAKFLHSKCC 298  
 DB 302 -NPVIYIMMKQF-----RNCMLTTICC 323

RESULT 4  
 US-08-820-521-2  
 ; Sequence 2, Application US/08820521  
 ; Patent No. 5942416  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bergsma, Derek  
 ; APPLICANT: Ganesh, Sathe  
 ; APPLICANT: Fuelterer, Wendy  
 ; APPLICANT: Mao, Joyce  
 ; TITLE OF INVENTION: CDNA CLONE HNFY20 THAT ENCODES  
 ; TITLE OF INVENTION: A NOVEL HUMAN 7-TRANSMEMBRANE RECEPTOR  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SmithKline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FASTSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/820,521  
 ; FILING DATE: 19-MAR-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Han, William T  
 ; REGISTRATION NUMBER: 34,344  
 ; REFERENCE/DOCKET NUMBER: GH50011  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 610-270-5219  
 ; TELEFAX: 610-270-4026  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 401 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-820-521-2  
 Query Match 8.1%; Score 122.5; DB 2; Length 401;  
 Best Local Similarity 24.0%; Pred. No. 7.3e-05;

Matches 71; Conservative 53; Mismatches 93; Indels 79; Gaps 18;  
 QY 10 FLAVIOLGIFNGIIVVNGIDILKHKMAPLDLISCLVSRIFLOLFIYVAVIY 69  
 DB 73 FSVYLLFELVGLPLN-LTALVAVFGKLRPVA-VDVYLLNLWLTASDILLLEFPRVEA 130  
 QY 70 I-----FFIEFMSCANCAILLFETNELMTWLGVCYCAKVASVRHPLFTLMKR--- 121  
 DB 131 ANGMHPLFELCPLSG--FFETFTIYL-TALFLAASIERFLSAHPL-WYTRPRLG 185  
 QY 122 ---ISKLVPMILG---SLVYSMIC--VFSKYAGFVWPYFLKFSQONATIQ---KED 170  
 DB 186 QAGIVSVACWILASAHCSVYVYIEFSGDISHQ-----GTNCTCYLEFRKD 231  
 QY 171 TLAT-----QIESFVAEFSPPLIFLFAVLLIFESLGR---HTRORNTVAGSRVGRGAP 223  
 DB 232 QLATLPLVLENAVLEVPVPLITSTCYSLVWILGRGSHRORR----- 277  
 QY 224 ISALL--SLISFLIY---FSH-----C-----MIKVLSSLKFHRRIFL 261  
 DB 278 VAGLALATILNPLVCGPYNVSHVGYIGCESPAMRIYVTLTSLNSGVDPFVYF 333

RESULT 5  
 US-08-748-506-22  
 ; Sequence 22, Application US/08748506  
 ; Patent No. 6159707  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ronnelt et al.  
 ; TITLE OF INVENTION: NOVEL SPERM RECEPTORS  
 ; NUMBER OF SEQUENCES: 31  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Leydig, Yoit & Mayer, Ltd.  
 ; STREET: Two Prudential Plaza, Suite 4900  
 ; CITY: Chicago  
 ; STATE: IL  
 ; COUNTRY: US  
 ; ZIP: 60601-6780  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/748,506  
 ; FILING DATE: 08-NOV-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/033,751  
 ; FILING DATE: 09-NOV-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; REFERENCE/DOCKET NUMBER: 74940  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-616-5700  
 ; TELEFAX: 312-616-5700  
 ; INFORMATION FOR SEQ ID NO: 22:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 327 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-748-506-22

Query Match 7.9%; Score 120.5; DB 3; Length 327;  
 Best Local Similarity 23.4%; Pred. No. 9.2e-05;  
 Matches 64; Conservative 48; Mismatches 110; Indels 51; Gaps 14;  
 QY 5 HT-IIFYFLAVIQFLGIFTNGIIVVNGIDILKHKMAPLDLIS-----CLAVSRIF 57  
 DB 28 HRLIFLFLHLAVLASLGMNLITITICVD--HRLQTPMVFILSMSSVCECETITVI 84

OY 58 LQLEFYVNVIVF---FIEFIMCSANCAILLFNELEMLATWLGVF---YCAKVASV 110  
 Db 85 POLL-----TIISGRKQIPFMACISOAFVLYVGATGFEL---LGVLSIDRFALICKPL 136  
 OY 111 RHPLEIMKMRISKVLPMWILGSLLYVS-MICVPHSKYAG-FMVPYFLRKFFSONATIQK 168  
 Db 137 HYPIMSPRMCFLLVYSLVGLFPMASPVYMLQSOFYCGPNIIHPFCD-EGPLANLSC 195  
 OY 169 EDTLAIQIFSVFAEFVSPLLEFAVLL-IFSLGHRTRQMRNTVAGSRVPGCAPISAL 227  
 Db 196 SETSIEMLEFTEL-----AIIVLFASLIAIRFVS-----NIVVTIVRLPSARERQRA- 243  
 OY 228 LSLSLFLLYFHSCHMIKVFSLSLKFIHRRIFEL 260  
 Db 244 -----FSTCSSHLIYLSIMYGSACAFIYL 266

RESULT 6  
 US-08-748-506-23  
 ; Sequence 23, Application US/08748506  
 ; Patent No. 6159707  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ronnett et al.  
 ; TITLE OF INVENTION: NOVEL SPERM RECEPTORS  
 ; NUMBER OF SEQUENCES: 31  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Leydig, Volt & Mayer, Ltd.  
 ; STREET: Two Prudential Plaza, Suite 4900  
 ; CITY: Chicago  
 ; STATE: IL  
 ; COUNTRY: US  
 ; ZIP: 60601-6780  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/748,506  
 ; FILING DATE: 08-NOV-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/033,751  
 ; FILING DATE: 09-NOV-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; REFERENCE/DOCKET NUMBER: 74940  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-616-5700  
 ; TELEFAX: 312-616-5600  
 ; INFORMATION FOR SEQ ID NO: 23:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 327 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-748-506-23

Query Match 7.9%; Score 120.5; DB 3; Length 327;  
 Best Local Similarity 23.4%; Pred. No. 9,2e-05;  
 Matches 64; Conservative 48; Mismatches 110; Indels 51; Gaps 14;  
 OY 5 HLIIYELLAVIOFLGFTNGCIIVVNGIDILKHKRMAPLDDLIS-----CLAVSRIF 57  
 Db 28 HURLIFELLHLAVLASLGMNLIITITVD---HRLQIPMFIFLSMSSVECCFTTIVT 84  
 OY 58 LQLEFYVNVIVF---FIEFIMCSANCAILLFNELEMLATWLGVF---YCAKVASV 110  
 Db 85 POLL-----TIISGRKQIPFMACISOAFVLYVGATGFEL---LGVLSIDRFALICKPL 136  
 OY 111 RHPLEIMKMRISKVLPMWILGSLLYVS-MICVPHSKYAG-FMVPYFLRKFFSONATIQK 168

Db 137 HYPIMSPRMCFLLVYSLVGLFPMASPVYMLQSOFYCGPNIIHPFCD-EGPLANLSC 195  
 OY 169 EDTLAIQIFSVFAEFVSPLLEFAVLL-IFSLGHRTRQMRNTVAGSRVPGCAPISAL 227  
 Db 196 SETSIEMLEFTEL-----AIIVLFASLIAIRFVS-----NIVVTIVRLPSARERQRA- 243  
 OY 228 LSLSLFLLYFHSCHMIKVFSLSLKFIHRRIFEL 260  
 Db 244 -----FSTCSSHLIYLSIMYGSACAFIYL 266

RESULT 7  
 US-08-785-928-1  
 ; Sequence 1, Application US/08785928  
 ; Patent No. 6087115  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gershengorn, Marvin C.  
 ; APPLICANT: Arvanitakis, Ieandros  
 ; APPLICANT: Geras-Raaka, Elizabeth  
 ; APPLICANT: Cesalman, Ethel  
 ; TITLE OF INVENTION: CONSTITUTIVELY ACTIVE G PROTEIN COUPLED  
 ; TITLE OF INVENTION: RECEPTOR OF HHV 8 AND METHOD OF IDENTIFYING NEGATIVE  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
 ; STREET: Clinton Square, P.O. Box 1051  
 ; CITY: Rochester  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 14603  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/785,928  
 ; FILING DATE: 22-JAN-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Goldman, Michael L.  
 ; REGISTRATION NUMBER: 30,727  
 ; REFERENCE/DOCKET NUMBER: 19603/1320  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (716) 263-1304  
 ; TELEFAX: (716) 263-1600  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 342 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-785-928-1

Query Match 7.9%; Score 120.5; DB 3; Length 342;  
 Best Local Similarity 22.4%; Pred. No. 9,7e-05;  
 Matches 59; Conservative 56; Mismatches 95; Indels 53; Gaps 14;  
 OY 11 LLAIVOPFLGFTNGCIIVVNGIDILKHKRMA-PLDLLSLAVSRIFLQLEFYVNVIV 69  
 Db 53 ILSL-FLIIVNGGLVYI---FCKHRSRAGAIDILLGICLNSICLSL--LAEVL 105  
 OY 70 IFFIEFIMCSANCAILLFNELEMLATWLGVFCAKVASVRHPLFI-----MLKMRISK 124  
 Db 106 MELPNIISTGCLCLEIFF-----YLVVYIDIFSVGVSLVRYLLAVSTRSNPK---KQ 158  
 OY 125 LVPMMILGSLIYVMT-----CVFHSKYAGFMVPYFLRKFFSONATIQKED-TLAIQIFS 178  
 Db 159 SLGAVLTSALLIALIYLSGDCACRHRSHVVD---FVSKQAMCYENAGMWTADWRLHVRTVS 215



QY 179 FVAEFSVPLILFVAVLLIFSLG----RHR-OMRNTVAGSRVPGAPISALLSLISF 233  
 Db 216 VTAGEFLPL-----ALLILFYALTWCYVRRTKIQARKKVG-----YIVAV 256  
 QY 234 LILYFSHCMIKVFSLSKFHRR 256  
 Db 257 VILFEVFCPPYHVLNLDLTLRR 279

RESULT 8  
 US-08-728-603-17  
 ; Sequence 17, Application US/08728603  
 ; Patent No. 6093806  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Caesarman, Ethel  
 ; APPLICANT: Knowles, Daniel M.  
 ; TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED  
 ; TITLE OF INVENTION: HERPESVIRUS  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP  
 ; STREET: Clinton Square, P.O. Box 1051  
 ; CITY: Rochester  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 14603  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/728, 603  
 ; FILING DATE: 10-OCT-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BRAMAN, SUSAN J.  
 ; REGISTRATION NUMBER: 34,103  
 ; REFERENCE/DOCKET NUMBER: 19603/720  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 716-263-1636  
 ; TELEFAX: 716-263-1600  
 ; INFORMATION FOR SEQ ID NO: 17:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 342 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-728-603-17

Query Match 7.9%; Score 120.5; DB 3; Length 342;  
 Best Local Similarity 22.4%; Pred. No. 9.7e-05;  
 Matches 59; Conservative 56; Mismatches 95; Indels 53; Gaps 14;

QY 11 LLAIVQFLCITFNGITVVVNGIDILKHKRMA-PDLLLSCLAVSRIFLOLFYFVNVIV 69  
 Db 53 ILSLI-FLINVLNGLVTYI-----FCKHRSRAGADILLGLICLSISL--LAEVL 105  
 QY 70 IFTEFTGNSCANCAILFELNELEMLATWLGVEYCAKVASVRHPLF-----WLKMRISK 124  
 Db 106 MFLFPIITISGCLREIF-----YLLVYIDIFSVCSLVRYLLAVYSRSPK--RQ 158  
 QY 125 LVPMMILGSLIYVSM-----CVFHSKYAGPMYPFLRKFSQNAITQKED-TLAIQIFS 178  
 Db 159 SLGWVLTSAALLIALVLSGDCRHSRVVD---PVSQAMCYENAGMTADWRLHVTVS 215  
 QY 179 FVAEFSVPLILFVAVLLIFSLG----RHR-OMRNTVAGSRVPGAPISALLSLISF 233  
 Db 216 VTAGEFLPL-----ALLILFYALTWCYVRRTKIQARKKVG-----YIVAV 256  
 QY 234 LILYFSHCMIKVFSLSKFHRR 256

Db 257 VILFEVFCPPYHVLNLDLTLRR 279

RESULT 9  
 US-08-928-692-12  
 ; Sequence 12, Application US/08928692  
 ; Patent No. 5958727  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brody, Howard  
 ; APPLICANT: Yaver, Deborah S.  
 ; APPLICANT: Lamsa, Michael  
 ; APPLICANT: Hansen, Kim  
 ; TITLE OF INVENTION: Methods for Modifying the Production of  
 ; TITLE OF INVENTION: a Polypeptide  
 ; NUMBER OF SEQUENCES: 80  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 59587270 No. 5958727disk of No. 5958727th America, Inc.  
 ; STREET: 405 Lexington Avenue  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 10174  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FASTSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/928, 692  
 ; FILING DATE: 12-SEPT-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lambiris, Elias J.  
 ; REGISTRATION NUMBER: 33,728  
 ; REFERENCE/DOCKET NUMBER: 4944, 200-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-867-0123  
 ; TELEFAX: 212-867-9655  
 ; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 524 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: No. 5958727e  
 ; US-08-928-692-12

Query Match 7.8%; Score 118.5; DB 2; Length 524;  
 Best Local Similarity 24.8%; Pred. No. 0.00026;  
 Matches 55; Conservative 30; Mismatches 64; Indels 73; Gaps 11;

QY 6 LIIYVLAIVQFLCITFNGI-----IVVYNGIDILKHKRMAPDLLLSCLAVSRIFLQ 59  
 Db 304 ILVALMLHVAQOFSI--NGIFYYSTSIPTAGIS-----KRYVATIGGAVANNVETA 354  
 QY 60 LFIFVYVNIIVIFTEFTGNSCANCAILFELNELEMLATWLGVEYCAKVASVRHPLFTMLK 119  
 Db 355 VSVFLVER-----AGRRSLFLIOMSGMFCA-----IFMSVG 386  
 QY 120 KRISLAVPMILGSLIYVSMICVFHSKYAGM-----VPEFL-RKFSQNAITQKEDTL 172  
 Db 387 LVLLNKFSWM-----SYVSMIAIF--LFVSFPEIGPGFIPFMVAEFSSQ---PRPAL 436  
 QY 173 AIQIFS-----FVAEFSVPLILFVAVLLIFSL 201  
 Db 437 AIAAASNMTCTNIVALCPQYINDFCGPYVFFLFAVLLAFTL 478

RESULT 10  
 US-08-466-103A-2  
 ; Sequence 2, Application US/08466103A

```

; Patent No. 5856124
; GENERAL INFORMATION:
; APPLICANT: Reppert, Steven M.
; APPLICANT: Edisawa, Takashi
; TITLE OF INVENTION: HIGH-AFFINITY MELATONIN
; TITLE OF INVENTION: RECEPTORS AND USES THEREOF
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,103A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/319,887
; FILING DATE: 07-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/261,857
; FILING DATE: 17-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00786/250002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 420 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; US-08-466-103A-2

```

```

Query Match 7.8%; Score 118; DB 2; Length 420;
Best Local Similarity 19.4%; Pred. No. 0.00022;
Matches 57; Conservative 64; Mismatches 119; Indels 54; Gaps 12;

```

```

OY 2 LESHLIYFLAVIOFLGFTNGIIVVNGIDILKRRKAPL-DLLSCLAVSRIFLQL 60
DB 31 LLSLAVLVLTIVDVGL---NILEY---LSVLNKKLQNGNLEVVSLSTADLVAV 83
OY 61 FIFVYVNVIVFTEFTMCSANCAILLEFNELEMLATWLGVCAYASVRHPLFTLMK 120
DB 84 YPVPVILIAIFQNGMTLGNHCQISGFLMGLSV-----IGSVNITAIAMINRCYICHL 138
OY 121 RISKLIV---PMWILSLVSMICVFHSKYAGFNV--PYELKRFESQNAATQKEDFLAI 174
DB 139 RYKLLNQRSTWYLGTLWTLLITIAIVPNEFVGSLOYDPRISCFEQAQ--TVSSSYTIV 196
OY 175 QIFSVAEFSVPLILFLEAVLLILFSLGRHTRQ-----MRTVA----- 213
DB 197 VVVHFLVPLSVYFECFL-RHWLVIVQKRVKRODFKQKLTQDLRNFELMFEVLEAVC 255
OY 214 -----GSAV---PGRAP-ISALLISFLILYFSHCKIRVFLSSLKFHTR 256
DB 256 WAPLNFGLAVAINPFAKIPENLFLVSYFMAYFNSCLNAINVIGVNLQNRK 309

```

```

RESULT 11
US-08-748-506-24

```

```

; Sequence 24, Application US/08/748506
; Patent No. 6159707
; GENERAL INFORMATION:
; APPLICANT: Ronnett et al.
; TITLE OF INVENTION: NOVEL SPERM RECEPTORS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Volt & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,506
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,751
; FILING DATE: 09-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 74940
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5600
; TELEFAX: 312-616-5700
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-748-506-24

```

```

Query Match 7.7%; Score 117.5; DB 3; Length 327;
Best Local Similarity 22.1%; Pred. No. 0.00019;
Matches 60; Conservative 51; Mismatches 112; Indels 49; Gaps 12;

```

```

OY 5 HL-IYFLLAVIOFLGFTNGIIVVNGIDILKRRKAPLIDLLS-----CLAVSRIF 57
DB 28 HRIILFELMLAVIASSMGNLITTYCVD---HRLQTPWTFLLSTFSVECCFTTVI 84
OY 58 IOLFIFVNVIVF---FIEFTMCSANCAILLEFNELEMLATWLGV-FYCAKVASVRHP 113
DB 85 POLT-----TILSGROKIPFWACFSQAFVVLGCAVAFVLMVAVSLDRFLAICKPLHY 139
OY 114 LEIWLKMRISKVPMIIGSLIYVS-MICVPHSKYAG-FMVPYELKRFESQNAATQKEDT 171
DB 140 TMSRMRCEFLVTVSLVGLFPMASPVVMLSQSFYCGENIIPHEFCO-FGPLANLSCSET 198
OY 172 LAIQTFSEVAESV---DLILFPAVLLILFSLGRHTRQMHNTVAAGSRVPGAPISALL 228
DB 199 RSIEMLFETLAIIVLFTSLLAIFAVSYIVTVIR-----LPSARERRA----- 243
OY 229 SILSLILYFSHCKIRVFLSSLKFHTRIFL 260
DB 244 -----FSTCSHLLIVLSLMTGSCVFIVL 266

```

```

RESULT 12
US-08-458-970A-11
; Sequence 11, Application US/08458970A
; Patent No. 5861272
; GENERAL INFORMATION:
; APPLICANT: Li, ET AL.
; TITLE OF INVENTION: C5a Receptor
; NUMBER OF SEQUENCES: 11

```

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
 ADDRESSEE: CECCHI, STEWART & OLSTEIN  
 STREET: 6 BECKER FARM ROAD  
 CITY: ROSELAND  
 STATE: NEW JERSEY  
 COUNTRY: USA  
 ZIP: 07068  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 INCH DISKETTE  
 COMPUTER: IBM PS/2  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: WORD PERFECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/458,970A  
 FILING DATE: June 2, 1995  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US94/09234  
 FILING DATE: 16 AUG 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MULLINS, J.G.  
 REGISTRATION NUMBER: 33,073  
 REFERENCE/DOCKET NUMBER: 325800-353  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-994-1700  
 TELEFAX: 201-994-1744  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 358 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-458-970A-11

Query Match 7.58; Score 113.5; DB 2; Length 358;  
 Best Local Similarity 21.0%; Pred. No. 0.00053;  
 Matches 69; Conservative 56; Mismatches 122; Indels 81; Gaps 15;

QY 5 HLIITFLAV---IOFLGIFNGIIVVNGIDLRKHKMAPDLLESLCLAVSRIFQLF 61  
 DB 23 HNYITVMIPTLSIIFVIGFNSLVAVIYIF-YKTKTVASVELNLNADLCFLPLP 81  
 QY 62 IYVVVIVYIEFIEF--IMCSANCAILLFNELEMLATWLGVEYCAKVASRHP----- 113  
 DB 82 IMAVVTAMEYRMPFGNYLCKIASASVSEFLYASVFLTCLSI--DRYLAVIHPKSRRLR 138  
 QY 114 -----LFTWLMKRSKLVPMMLISLLYVS-----MICVFH----- 144  
 DB 139 RMLVAKVTCIITWLAGIASL-PALITHRNVEFIENTNITVCAFIYESONSTLPGLGLT 197  
 QY 145 SKYAGMVPYF-----LRFESONATIQED-----TLAIQFSVAEFS-VP 186  
 DB 198 KNIGLGLFELIILTSYTLIMKAKKAYEIOKNKPRNDIIFKIMAVLFEF---FSWIP 254  
 QY 187 LLIETFAVLLIFSLGRHROMRNTYAGSRVGRGAPISALLSILSLIYSHCMIKYF 246  
 DB 255 HQIETFLDYIIOGIRDRRIA--DIVDAMP-----TTCIAYFNCLNPLF 300  
 QY 247 LSLKFIHRIEFLFVFLVIGIYPSGHS 274  
 DB 301 YGFLGKKFKRY-FLQLIKTIYPPAKSHS 327

RESULT 13  
 US-08-118-270-65  
 ; Sequence 65, Application US/08118270  
 ; Patent No. 5508384  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Murphy, Randall B.  
 ; APPLICANT: Schuster, David I.

TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
 TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
 NUMBER OF SEQUENCES: 348  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BROWDY AND NEIMARK  
 STREET: 419 Seventh Street, N.W., Suite 300  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/118, 270  
 FILING DATE: 09-SEP-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/943,236  
 FILING DATE: 10-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Townsend, Kevin G.  
 REGISTRATION NUMBER: 34,033  
 REFERENCE/DOCKET NUMBER: MURPHY-2A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-628-5197  
 TELEFAX: 202-737-3528  
 TELEX: 248633  
 INFORMATION FOR SEQ ID NO: 65:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 286 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-118-270-65

Query Match 7.08; Score 107; DB 1; Length 286;  
 Best Local Similarity 21.68; Pred. No. 0.0019;  
 Matches 66; Conservative 52; Mismatches 131; Indels 56; Gaps 11;

QY 8 IYFLAVIOTFLGIFNGIIVVNGIDLRKHKMAPDLLESLCLAVSRIFQLFITYVNV 67  
 DB 1 LFLFLSLXVYLVTENMLIIRNPTLHKPY-----YFLFLFIWYVTV 47  
 QY 68 IVIFIEFIMCSANCAILL--FINELMLATWLGVEYCAKVA-----SVRHPLE-- 115  
 DB 48 TTPKMGFISGRKHGOISFFACMTOLYFPLGLGCTECVLLAMAVDRYVAICHPLHP 107  
 QY 116 IMLKMRISKLVPMMLISLLYVSMICVFSKYAGMVPYFLRKFESONATIQEDTLAIQ 175  
 DB 108 VIVSRITVLGSM---AGFGISMKVVFLLISRLSYCGPNTFNHFFC-----DVSPL 156  
 QY 176 IFSEVAEFSVPLLFPLAVALLISLGRHROMRNTYAGS---RVP---GRGAPISALLS 229  
 DB 157 NLSCDTMSTAEITDFVIAIFILGLP-----SVYGASVMTIPSAAGHKAFTSCAS 207  
 QY 230 ILSFLIFESOMIKVFLFSLKFIHRIEFLFVIGIYPSGHSLLILITLGNPKLRON 287  
 DB 208 HLTAVYITFYA---ASIFIYARPKALSFTDNKLVSLAVAVLPLEPNPIYICLRNODVKRA 264  
 QY 288 AKKFL 292  
 DB 265 LKRTL 269

RESULT 14  
 PCT-US93-08528-65  
 ; Sequence 65, Application PC/TUS9308528  
 ; GENERAL INFORMATION:  
 ; APPLICANT: New York University

```

; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF INVENTION: 348
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 286 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08528-65

```

```

Query Match 7.0%; Score 107; DB 4; Length 286;
Best Local Similarity 21.6%; Pred. No. 0.0019;
Matches 66; Conservative 52; Mismatches 131; Indels 56; Gaps 11;

QY 8 IYFLAVIOFLGFTNGIIVVNGIDILKHKMAPDLDSCLAVSRIFLOLFYVNV 67
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 LFLFLSLXVVLVTEMLTIIRNHPILKPM-----YFLFLLEIYVTV 47
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 IVYFFIEFMCANCAITL--FTNELEMLATWLGVEYCAKVA-----SVRHPLF-- 115
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 48 TIRPLMGFISKEHNGQLISFCAMTQLYFFELGCEVLLAVMAYDRVVAICHPILHP 107
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 IMLKMRISKLVPMWILGSLVSMICVFNHKGAFVYFLKRFESQNAITQKEDFLAIQ 175
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 108 VIYSSRIZVIGSM---AGRCISMVAVFLISRLSYCGPNTINHFC-----DIVSPLL 156
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 IESFVAEFSVLLIFLFAVLLIFSLGRHTRQMRNTVAGS---RVP---GRGAPISALLS 229
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 157 NLSCTDMSTAEIADDEVIATPILGPL-----SVTASVYKRISAGAKHAEFTCAS 207
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 ILSEFLIFESHCHIKVFLSLKAFHTRFT--FLFELVIGIYPSGSLILIGNPKLON 287
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 208 HLTVVLIIFVA---ASIFIVARPKALSAFTDNKLVAVLAVIYFLNPITCYLRNQVKRA 264
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 AKRFL 292
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 265 LRRTL 269

```

```

RESULT 15
US-08-148-209A-4
; Sequence 4, Application US/08148209A
; Patent No. 5556780
; GENERAL INFORMATION:

```

```

; APPLICANT: Dzau, Victor J
; APPLICANT: Mukoyama, Masashi
; TITLE OF INVENTION: TYPE-2 ANGIOTENSIN II RECEPTOR AND GENE
; NUMBER OF INVENTION: 16
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,209A
; FILING DATE: 05-NOV-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58491-1/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-148-209A-4

```

```

Query Match 6.9%; Score 104.5; DB 1; Length 359;
Best Local Similarity 21.3%; Pred. No. 0.0044;
Matches 70; Conservative 58; Mismatches 118; Indels 83; Gaps 17;

QY 5 HLIFVLLAV---IOFLGIFTNGIIVVNGIDILKHKMAPDLDSCLAVSRIFLOLF 61
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 24 HSYIFEMIPTLSIIFVGVIFGNSLVIVYIF-YMKLTVASVFLNLALADLCFLTLTP 82
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 IYVNVYVIEFIEF--IMCSANCAITLFTNELEMLATWLGVEYCAKVASRHP----- 113
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 83 LMAVVTAMETRWPFQGNHLCKLASASVSFNLVASVFLTLCLST---DRLATVHPKSRRL 139
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 -----LFTWLKMRISKLVPMWILGSLVYS---MICEFH----- 144
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 RMLVAKVTCITIIWMAGIASL-PAVIRHNVYFIENITVCAFIYESRNTLPIGLTL 198
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 145 SKYAGFNPYF-----LAKFESQNAITQKED-----TLAIOESFAEFS-VP 186
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 199 KNILDFLEPFLIILTSYLIMKALKAKAYEIQKNRPRNDIPRIIMAYLFFEF---FSWVP 255
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 LLIPLFAVLLIFSLG-RHTRQMRNTVAGSRVPGAPISALLSILSLIFLFSHCHIKV 245
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 256 HQIFTF--LDVLQIGVHDKKID-IYDTAMP-----ITICIAVFNNCLNPL 300
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 FLSSLKFIHRFIFLFLVIGIYPSGHS 274
   | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 301 FYGFLGKKKKRY-FLQLLKXIYIPRAKSHS 328

```

```

Search completed: March 15, 2001, 12:57:20
Job time: 4696 sec

```



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: March 15, 2001, 12:32:39 ; Search time 76.8 Seconds  
(without alignments)  
133.124 Million cell updates/sec

Title: US-09-510-332-1  
Perfect score: 1521  
Sequence: 1 MLESHLLIYFLAVIQFLG.....GNPKIKONAKFLHSKCCQ 299

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_36.\*  
1: /cgn2\_2/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /cgn2\_2/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /cgn2\_2/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /cgn2\_2/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /cgn2\_2/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /cgn2\_2/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /cgn2\_2/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /cgn2\_2/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /cgn2\_2/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /cgn2\_2/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /cgn2\_2/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /cgn2\_2/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /cgn2\_2/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /cgn2\_2/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /cgn2\_2/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /cgn2\_2/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /cgn2\_2/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /cgn2\_2/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /cgn2\_2/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /cgn2\_2/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	263.5	17.3	256	Y42388	Amino acid sequenc
2	124.5	8.2	348	R38483	Rhodopsin protein.
3	124.5	8.2	348	R38483	Rhodopsin. Homo s
4	123.5	8.1	309	R48735	G-protein coupled
5	123.5	8.1	309	W02707	G-protein coupled
6	122.5	8.1	401	W59924	Human 7-transmembr
7	120.5	7.9	342	W37976	Kaposi's sarcoma a
8	118	7.8	354	W25926	Xenopus melatonin
9	118	7.8	354	W25927	Xenopus melatonin
10	117.5	7.8	420	R88409	High-affinity mela
11	117.5	7.7	310	W19105	Rat pheromone rece
12	117.5	7.7	522	W17836	Rat glucose transp

13	117.5	7.7	690	Y13352	Amino acid sequenc
14	117.5	7.7	690	Y83225	PRO288 Polypeptide
15	116	7.6	349	P90554	Bovine rhodopsin.
16	114	7.5	123	Y27593	Human secreted pro
17	112.5	7.4	354	Y57086	Rhodopsin amino ac
18	112.5	7.4	476	Y70157	Staphylococcus aur
19	109.5	7.2	362	Y16630	Human putative Adr
20	109.5	7.2	389	Y19909	B. burgdorferi ant
21	109.5	7.2	449	Y19908	B. burgdorferi ant
22	108	7.1	337	Y19563	Human G protein co
23	107.5	7.1	312	R27875	Oodant receptor c
24	107	7.0	286	R48744	G-protein coupled
25	107	7.0	286	W02716	G-protein coupled
26	106.5	7.0	372	W86323	Kidney injury asso
27	104	6.8	342	R25834	Guinea pig platele
28	103.5	6.8	327	W21666	Rat spermatoz chem
29	103.5	6.8	437	Y34786	Amino acid sequenc
30	103.5	6.8	690	Y79383	Human G protein co
31	103	6.8	337	W85047	Rat pheromone rece
32	103	6.8	337	W85047	A human 7-transmem
33	103	6.8	337	W75799	Human 7-transmembr
34	103	6.8	342	W12243	Guinea pig platele
35	103	6.8	380	R53750	Seven transmembran
36	103	6.8	380	W48731	Human R20 seven tr
37	103	6.8	380	Y24952	G-protein-coupled
38	103	6.8	380	Y84442	G-protein-coupled
39	103	6.8	380	Y87465	Human APJ receptor
40	102	6.7	780	Y41089	Human lectioned-in-1
41	102	6.7	1021	Y77728	Human G protein-co
42	102	6.7	1114	Y41092	Peptide Seq ID No:
43	102	6.7	1123	Y41086	Human lectioned-in-1
44	102	6.7	1177	Y41085	Human lectioned-in-1
45	102	6.7	1403	Y41087	Human lectioned-in-1

## ALIGNMENTS

RESULT	1	ALIGNMENTS
Y42388	Y42388 standard. Protein; 256 AA.	
XX	Y42388;	
AC	09-DEC-1999 (first entry)	
XX		
DE	Amino acid sequence of p127-1.	
XX		
KW	secreted protein; cDNA library; clone; transmembrane protein;	
KW	signal sequence cloning; hybridization cloning; gene therapy;	
KW	receptor.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	Peptide	8..20
FT	Protein	/label= Leader/Signal peptide
FT		21..256
FT	Misc-difference	/label= Mature protein
FT		64
FT		/label= unknown
FT		/note= "encoded by twa"
XX	W09942470-A1.	
XX		
PN	26-AUG-1999.	
XX		
PD		
XX	18-FEB-1999;	99WO-US03458.
PF		
XX	18-FEB-1998;	98US-0075038.
PR	17-FEB-1999;	99US-0251600.
XX		
PA	(GENY ) GENETICS INST INC.	

XX Jacobs K, McCoy JM, Lavelle ER, Collins-Racie LA, Merberg D;  
 PI Treacy M, Agostino MJ, Steinger RJ;  
 XX WPI: 1999-518580/43.  
 DR N-PSDB: 220861.  
 XX  
 PT New polynucleotides encoding human secreted proteins used for  
 XX therapeutic, diagnostic and research purposes.  
 PS Claim 27; Page 114-115: 125pp; English.

XX This is the amino acid sequence of the p127-1 protein, which is  
 CC derived from the p127-1 clone isolated from a human adult blood  
 CC (lymphoblastic leukemia MOLT-4) CDNA library.  
 CC The pns and proteins of the invention are predicted to have biological  
 CC activities which would make them suitable for treating, preventing or  
 CC ameliorating medical conditions in humans and animals, although no  
 CC supporting data is given. Suggested activities include nutritional  
 CC activity, cytokine and cell proliferation/differentiation activity,  
 CC immune stimulating (e.g. as vaccines) or suppressing activity,  
 CC hematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic  
 CC and thrombotic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumor invasion suppressor activity, and tumor  
 CC inhibition activity. The pns are also stated to be useful for gene  
 CC therapy. Other activities include inhibiting the growth, infection or  
 CC function of bacteria, fungi, viruses and other parasites; effecting  
 CC bodily characteristics such as, e.g. weight, color, skin, etc.;  
 CC effecting biorhythms or cardiac cycles; enhancing fertility; treatment  
 CC of depression; treatment of pain; hormonal or endocrine activity.  
 XX Sequence 256 AA;

Query Match 17.3%; Score 263.5; DB 20; Length 256;  
 Best Local Similarity 28.8%; Pred. No. 7.9e-23;  
 Matches 74; Conservative 58; Mismatches 74; Indels 51; Gaps 8;

QY 7 IIVFLAVIOFLGIFTNGITVVNGIDLIKRRKMAPLDLLSCLAVSRIFQLFFYYN 66  
 Db 7 IISLLVVFVIGNFANGFIVNSLEWKQKISFADQLTALAVSYGL-----LW 60  
 QY 67 VIVVFEEFIMCSANCAILFT-----NELEMLATWLCVFYCAKAVSRHPLFW 117  
 Db 61 VILXHYAEVLNPGSYSLGVRLCTINAWVCNHSIWALESILFYILIKANFSNFIHL 120  
 QY 118 LKMRISKLVPMILGSLLYVSMIC-----VPHSKYAGFNVPRFLRKFSSQATIQ 167  
 Db 121 Ikrriksvplvlllgellf--lvchlvvnmde smwtkyegnv-----swelkls 169  
 QY 168 KEDTLAIQIFSEVAEFSVPLILFLFVALLIFSLGRHTRQMRTVAGSRVPGGAP--- 223  
 Db 170 dphldmctvltlanl-1pftclslslflllclslckhikmm-----qfhgyspsdnt 221  
 QY 224 ---ISALSLISFLICY 237  
 Db 222 kvhkalqvtcsflllf 238

RESULT 2  
 R38483  
 ID R38483 standard; Protein: 348 AA.  
 XX R38483;  
 AC  
 XX  
 DT 11-NOV-1993 (first entry)  
 XX  
 DE Rhodopsin protein.  
 XX  
 KW Human; rhodopsin; mutant; retinal degeneration; primer; probe;  
 KW hereditary.  
 XX

OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FH Misc-difference 249  
 FT /note= "sequence unclear at this position"  
 XX

XX MO9312134-A.  
 XX  
 XX 24-JUN-1993.  
 XX  
 XX 08-DEC-1992; 92WC-US10536.  
 XX  
 XX 11-DEC-1991; 91US-0805123.  
 XX

XX (HARD ) HARVARD COLLEGE.  
 XX Berson EL, Dryja TP;  
 XX WPI: 1993-214088/26.  
 XX N-PSDB: Q43543.  
 XX

XX Probe or primer contg. sequence of human retinal degeneration  
 PT slow protein mutant - used to diagnose hereditary retinal  
 PT degenerative diseases  
 XX  
 XX Disclosure: Fig 1; 56pp; English.

XX The sequence given represents the human rhodopsin protein. Mutant  
 CC rhodopsin proteins cause retinal degeneration. The DNA encoding the  
 CC mutant DNA sequences may be identified using primers/probes described  
 CC in the invention (see also Q43545-48) and may be used to diagnose  
 CC hereditary retinal degeneration.  
 XX

Sequence 348 AA;

Query Match 8.2%; Score 124.5; DB 14; Length 348;  
 Best Local Similarity 20.8%; Pred. No. 2.1e-06;  
 Matches 68; Conservative 63; Mismatches 125; Indels 71; Gaps 16;

QY 1 MESHUIITFLAVIOFLGIFTNGITVVNGIDLIKRRKMAPLDLLSCLAVSRIFQL 59  
 Db 39 mlaaym--flllvigfpinfillyvtv-----qhkrlrplnylllnlavadtfw 87  
 QY 60 LFIYVNVIVIFIEFIMCSANCAILFT-----NELEMLATWLCVFYCAKAVSR 111  
 Db 88 lggfcscltyslhyfvgfpgtcnlegfcltggelalvslvdleryvvvckpmsnfr 147  
 QY 112 ---HPL---FIW---LKMRIKSLVPMILGSLLYVSMICVFSKSYAGFNVPRFLR--- 157  
 Db 148 fgenhalingvafvwmalacaaplagw-----ary-----lpeglgcsc 187  
 QY 158 --KFSQONATIQKEDTLAIQIFSEVAEFSVPLILFLFVALLIFSLGRHTRQMRTVAGS 215  
 Db 168 gldyylkpevne---stvlmrvvhflipmlilfctcygqvlvtvkaeaaqgges-alc 243  
 QY 216 RVPGRGAPISALLISLFLILYFSCHMIKVFSLKFIHRRIFL-----FFITLVGIYPS 271  
 Db 244 gkaekxvtlrmvliwlaflfcwpyasavafylfthgsnfgplfncltpafefkasaaly-- 301  
 QY 272 GHSLLILGNPKLKONAKKFLHSHKCC 298  
 Db 302 -npvilylmmnkqf---trncmltllic 323

RESULT 3  
 R93116  
 ID R93116 standard; Protein: 348 AA.  
 XX R93116;  
 AC  
 XX  
 DT 06-JUL-1996 (first entry)  
 XX



```

XX Rhodopsin.
XX Human; rhodopsin; transversion; mutation; retinitis pigmentosa;
XX probe; primer; hybridisation; polymerase chain reaction; PCR;
XX eye; rod; retina; diagnostic; prenatal diagnosis; photoreceptor.
XX
XX Homo sapiens.
XX
FH Key location/Qualifiers
FT Misc-difference 23 /note= "Mutated to His in retinitis pigmentosa"
FT
XX US5498521-A.
XX
XX 12-MAR-1996.
XX
XX 24-JAN-1990; 90US-0469215.
XX
XX 11-MAR-1993; 93US-0033081.
XX 24-JAN-1990; 90US-0469215.
XX 11-DEC-1991; 91US-0805123.
XX
XX (HARD ) HARVARD COLLEGE.
XX
XX Berson EL, Dryja TP;
XX
XX MPI: 1996-159684/16.
XX N-PSDB: T17115, T17116.
XX
XX Diagnosis of hereditary retinal degenerative diseases e.g. retinitis
XX pigmentosa, - caused by a human photoreceptor protein mutation, by
XX detection of the mutation by PCR amplification or hybridisation
XX methods
XX
XX Example 1; Column 19-24; 71pp; English.
XX
XX This sequence represents human rhodopsin, and is encoded by
XX sequences T17115 (without introns) and T17116 (with introns).
XX Substitution of histidine for the normal nonpolar amino acid
XX proline at position 23 results in a dysfunctional or absent
XX molecule, affecting rod function, and is linked with autosomal
XX dominant retinitis pigmentosa. Probes and primers specific for
XX this mutation may be used as diagnostic agents. Mutations in the
XX retinal degeneration slow protein and retinal rod
XX CGMP-phosphodiesterase genes are also implicated in retinitis
XX pigmentosa. Detection of any of these mutations in a foetus or
XX patient may be used in diagnosis.
XX
XX Sequence 348 AA:
XX
XX Query Match 8.2%; Score 124.5; DB 17; Length 348;
XX Best Local Similarity 20.7%; Pred. No. 2,1e-06;
XX Matches 68; Conservative 64; Mismatches 125; Indels 71; Gaps 16;
XX
XX 1 MESHLLIYELLAVIQFLGIFNGIIVVNGIDLIKRRK-APDDLISCAVSRIFIQ 59
XX ||::: ||||: |::: |::: |::: |::: |::: |::: |::: |::: |:::
XX Db 39 mlaaym---fllylgfplnfllytvc-----qhkklrtpnylllnlavadlfmw 87
XX
XX 60 LFFFYVNVIVPIETETIKMSANCAILLFT---NELELW---LAWLWLFVFCARVASYR 111
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 88 lggftcltlyslhytvtfgptgcnglegfcltjggetalsvllateryvvvckpmsntr 147
XX
XX 112 ----HBL----FIW----LKMRIKSLVPMMLIGSLVSMICVFHSRYAGFMVPLYR--- 157
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
XX Db 148 fgenhaimgvafcwmalacaaaprlagw-----sry-----lpeglgqsc 187
XX
XX 158 --KFSQNAATIQEDPLAIQIFSFVAEFSVPLLIPLFAVLLIFSGLRHTROMKRNVAQS 215
XX ::::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
XX Db 188 gldyultkpevne---sfvlymfvvhftlpmlliffcygqlvtvkeaaagqes-att 243
XX
XX 216 RVGRGAPISALLSIISFLLYFSHCKIKVFLSLKFMHRRFLFEL---FFILVIGIYPS 271
XX ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

```

Db	244	gkaeekrtmvinwiafllwcpyaasvafiflhgsgntfplmtlpatfaksaaly--	301
Qy	272	GHSLLILGNPKLKNAMKFFLLHSSCCQ	299
		: :   :   : : : :   :   :	
Db	302	-npvlyimmkqf----fncmltlccr	324
RESULT	4		
ID	R48735	standard; Protein; 309 AA.	
XX	R48735		
XX	R48735;		
XX	07-JUN-1996	(first entry)	
XX			
DE	G-protein coupled human rhodopsin protein.		
XX			
KW	G-protein coupled receptor; ligand binding assay; transmembrane domain;		
KW	psychotic disorder; schizophrenia; dopamine; GMP; adenosine; thrombin;		
KW	muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;		
KW	rhodopsin; opsin; odorant; cytomagalovirus.		
XX			
OS	Homo sapiens.		
XX			
PN	MO9405695-A1.		
PD	17-MAR-1994.		
XX			
PF	09-SEP-1993; 93WO-US08528.		
XX			
PR	10-SEP-1992; 92US-0943236.		
XX			
PA	(UYMY ) UNIV NEW YORK STATE.		
XX			
PI	Murphy RB, Schuster DI;		
XX			
DR	WPI; 1994-101120/12.		
XX			
PT	polypeptides of G-coupled receptor proteins (GPRs) - useful for		
PT	binding GPR ligands or modulating GPR binding		
XX			
XX	Disclosure: Page 111-112; 160pp; English.		
PS			
CC	Proteins R48685-R48758 represent a range of G-protein coupled receptor		
CC	proteins selected from GMP, adenosine, muscarinic acetylcholine,		
CC	adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,		
CC	odorant, cytomagalovirus and other G-protein coupled receptors. The		
CC	receptor proteins were used to design polypeptides, pref. based on the		
CC	transmembrane domains, for use in G-protein coupled receptor ligand		
CC	binding assays. The polypeptide fragments retain biological activity		
CC	such as binding a GPR ligand or modulating GPR ligand binding to a GPR		
CC	(see R48759-R48758, R50569-R50807 and R89189-R89195 for examples of		
CC	polypeptide fragments). The polypeptide fragments can be used in		
CC	compositions for treating subjects suffering from a pathology related to		
CC	a GPR abnormality e.g. a psychotic disorder such as schizophrenia.		
XX			
SQ	Sequence 309 AA:		
	Query Match	8.1%; Score 123.5; DB 15; Length 309;	
	Best Local Similarity	21.1%; Pred. No. 2,3e-06;	
	Matches 69; Conservative 64; Mismatches 121; Indels 73; Gaps	17,	
Qy	1	MESHSLIYFLAVIOFLGIFFTNGIIVVNGIDLIKRRK-APDDLISCLAVSRIFIQ	59
		: : : :   :   : : : :   :   :	
Db	2	mlaaym---fllylglfplnfllyltv-----ghkkitrplylllnlavadlfnw	50
Qy	60	LFIIFYNVIVIFIEETMCSANCAILLFINEL-----ELMLATWLGIFYCAKVASVR-	111
		: : : :   :   : : : :   :   :	
Db	51	lqgftstltytslmglyivfgptcgnleqfialtylgelalslvalateryyyvckpmsnff	110
		: : : :   :   : : : :   :   :	
Qy	112	---HPL-----FIWLKRRKSLVP-----WMILGSLLYVSMICVFSKRYAGFMVPEFLR---	157
		: : : :   :   : : : :   :   :	

```

Db 111 genhalmgvaftw-malacaaplagw-----sry-----ipeqlqsc 149
Oy 158 --KFSQGNATIQKEDLAIQISFVAEFSVPLILFPAVLLIFSLGRTRQMRNTVAGS 215
Db 150 gldyyltkpvevne--sflymfvhftlplilf-fcygqlvftvkeaaqqes-att 204
Oy 216 RVPGRGAPISALSLISFLILFESHCMIKVFLSLKFHIRFTFL---FELIVIGIYPS 271
Db 205 qkaekvtrmvlmvaflfcwpyasvafyflftngsnfpglfnlpaftaksaaly-- 262
Oy 272 GHSLLILGNPKLONAKKFLHLSKCC 298
Db 263 -npvlylfnkqf---rncmlqlqcc 284

RESULT 5
W02707
ID W02707 standard; peptide: 309 AA.
XX
AC W02707;
XX
DE 13-NOV-1996 (first entry)
XX
DE G-protein coupled human rhodopsin.
XX
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW odorant; cytomagalovirus; serotonergic.
XX
XX Homo sapiens.
XX
XX US5508384-A.
XX
XX 16-APR-1996.
XX
XX 10-SEP-1992; 92US-0943236.
XX
XX 09-SEP-1993; 93US-0118270.
XX
XX 10-SEP-1992; 92US-0943236.
XX
XX (UYNV ) UNTV NEW YORK STATE.
XX
XX PA
XX PI Murphy RB, Schuster DI;
XX
XX WPI: 1996-208785/21.
XX
XX New dopamine receptor peptide - useful as antipsychotic agent, e.g.
XX for treating schizophrenia
XX
XX Disclosure: Column 139-140; 184pp; English.
XX
XX Proteins W02657-W02730 represent a range of G-protein coupled receptor
XX (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
XX adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
XX odorant, cytomagaloviral and other GPR proteins. The receptor proteins
XX were used to design polypeptides, pref. based on the transmembrane
XX domains, for use in G-protein coupled receptor ligand binding assays.
XX The polypeptide fragments retain biological activity such as binding a
XX GPR ligand or modulating GPR ligand binding to a GPR (see W02747-W02999
XX for examples of polypeptide fragments). The polypeptide fragments can
XX be used in compositions for treating subjects suffering from a pathology
XX related to a GPR abnormality e.g. a psychotic disorder such as
XX schizophrenia.
XX
XX Sequence 309 AA;

```

```

Query Match 8.1%; Score 123.5; DB 17; Length 309;
Best Local Similarity 21.1%; Pred. No. 2.3e-06;
Matches 69; Conservative 64; Mismatches 121; Indels 73; Gaps 17;
Oy 1 MLESLLIFFLAVIQFLGFTNGIIVVVGIDILIRKM-APDLLLSCLAVSRIFLQ 59

```

```

Db 2 mlaaym--flilvgfplnflilyvtv-----ghkkrlrplyllnlavadifmw 50
Oy 60 LFIIFYNVNVIIFEFIFMCSANCAILLFTINEL-----ELMLATWLGVEYCAKVASVR- 111
Db 51 lsgftctlytslnhyfvfpgtgnlqgfafatlggelaalslwaleryvvvcgpmnftf 110
Oy 112 ---HPL---FTWLKMRISKLPV---WMILGSLVYSNICVFSKRYAGFMVYFLR--- 157
Db 111 genhalmgvaftw-malacaaplagw-----sry-----ipeqlqsc 149
Oy 158 --KFSQGNATIQKEDLAIQISFVAEFSVPLILFPAVLLIFSLGRTRQMRNTVAGS 215
Db 150 gldyyltkpvevne--sflymfvhftlplilf-fcygqlvftvkeaaqqes-att 204
Oy 216 RVPGRGAPISALSLISFLILFESHCMIKVFLSLKFHIRFTFL---FELIVIGIYPS 271
Db 205 qkaekvtrmvlmvaflfcwpyasvafyflftngsnfpglfnlpaftaksaaly-- 262
Oy 272 GHSLLILGNPKLONAKKFLHLSKCC 298
Db 263 -npvlylfnkqf---rncmlqlqcc 284

```

```

RESULT 6
W59924
ID W59924 standard; protein: 401 AA.
XX
XX W59924;
XX
XX 07-DEC-1998 (first entry)
XX
XX Human 7-transmembrane receptor HNFEDY20.
XX
XX HNFEDY20; G-protein coupled receptor; human; infection; HIV; pain;
XX cancer; anorexia; asthma; Parkinson's disease; acute heart failure;
XX hypotension; hypertension; urinary retention; osteoporosis;
XX angina pectoris; myocardial infarction; ulcer; allergy;
XX benign prostatic hypertrophy; psychosis; anxiety; schizophrenia;
XX manic depression; delirium; dementia; mental retardation;
XX dyskinesia; Huntington's disease; Gilles de la Tourette's syndrome;
XX therapy; diagnosis.
XX
XX Homo sapiens.
XX
XX EP866126-A1.
XX
XX 23-SEP-1998.
XX
XX 16-FEB-1998; 98EP-0301122.
XX
XX 19-MAR-1997; 97US-0820521.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX
XX Bergsma DJ, Fuechterer WS, Mao JY, Sathe GM;
XX
XX WPI: 1998-482962/42.
XX
XX N-P.SDB: V53631.
XX
XX New polynucleotides and polypeptides encoding a novel human
XX 7-transmembrane receptor - useful for diagnosing and treating e.g.
XX cancer, osteoporosis and Parkinson's disease and infections caused
XX by HIV-1 or -2.
XX
XX Claim 1; Page 18-19; 24pp; English.
XX
XX This polypeptide comprises HNFEDY20, a novel human 7-transmembrane
XX G-protein coupled receptor that shows about 30.8% identity in 299
XX amino acid residues with the thrombin receptor. Its amino acid
XX sequence was deduced from an isolated HNFEDY20 polynucleotide
XX sequence (see V53631). The invention relates to HNFEDY20
XX polypeptides and recombinant materials and methods for their

```



```

Query Match 7.88; Score 118; DB 18; Length 354;
Best Local Similarity 19.48; Pred. NO. 1.2e-05;
Matches 57; Conservative 64; Mismatches 119; Indels 54; Gaps 12.

OY 2 LESHLLIYFPLAIOELGIFNGIIIVVANGIDIKRHKMPL-DLLSLCLANSRIFDL 60
    | | | : : : | | : : : : : | : : : : :
Db 31 ltsalavvllftlvadvg-----nllyl---lsvlrnkklqmagllfvsladlvav 83
    | | | : : : | | : : : : : | : : : : :

OY 61 FIFYVNVVIFIEFIEMCSANCAILLFNELEMLATMLGVCYKAVASVRHPLFLIMKM 120
    : : | : | : : : | : | : : : : : | : : : : :
Db 84 ypyrvllallfngwvlglnlhbqisgflmgsv-----lgsvfrlttatalnyrychsl 138
    | | | : : : | : : : : : | : | : : : : :

OY 121 RISKLV---PMMLIGSLLYVSMICVFRHSKYAGMV--PFLRKFFSONATIQKEPTLAI 174
    | | | : : : | : : : : : | : | : : : : :
Db 139 rydkllyngstscwcyglltwlltlltlaivpntfvsglqydpri facteq--tvsssytlcv 196
    | | | : : : | : : : : : | : | : : : : :

```

QY	175	QISFVAEFSVPLILFLFAVLLILFSLGHTNQ-----MNNTVA-----	213
Db	197	vvhflvplsvfcyl-ftwlvlgvkhrrvqddfkqklgtcdlnfltmfvevylfavc	255
OY	214	-----GSRV---PGRGAP-ISALSLISPLILFVFSCHQIKVFLSSLKPHRR	256
Db	256	waplnflglavlnphvcpkfpewlftvlsyftmayfnsclnavlygvlnqnfirk	309
RESULT	9		
ID	W25927	standard; Protein: 354 AA.	
XX	W25927;		
XX	10-NOV-1997	(first entry)	
XX			
DE	Xenopus melatonin receptor MEK-1ab.		
XX			
KM	Allele: Xenopus laevis; melatonin; receptor; untranslated region; PCR;		
KM	mRNA; half-life; skin; amplification; primer; polymerase chain reaction;		
KM	transmembrane domain; cellular signaling; inhibitor; adenylyl cyclase;		
KM	modulation; intracellular; cyclic GMP; inhibitor; phosphodiesterase.		
XX			
OS	Xenopus laevis.		
XX			
FH	Key	Location/Qualifiers	
FT	Domain	1..32	
FT		/note= "Extracellular N-terminal domain"	
FT	Domain	33..56	
FT		/note= "transmembrane domain 1"	
FT	Domain	57..68	
FT		/note= "intracellular domain 1"	
FT	Domain	69..92	
FT		/note= "transmembrane domain 2"	
FT	Domain	93..107	
FT		/note= "extracellular domain 2"	
FT	Domain	108..125	
FT		/note= "transmembrane domain 3"	
FT	Domain	126..151	
FT		/note= "intracellular domain 2"	
FT	Domain	152..171	
FT		/note= "transmembrane domain 4"	
FT	Domain	172..193	
FT		/note= "extracellular domain 3"	
FT	Domain	194..219	
FT		/note= "transmembrane domain 5"	
FT	Domain	220..243	
FT		/note= "intracellular domain 3"	
FT	Domain	244..268	
FT		/note= "transmembrane domain 6"	
FT	Domain	269..280	
FT		/note= "extracellular domain 4"	
FT	Domain	281..301	
FT		/note= "transmembrane domain 7"	
FT	Domain	302..354	
FT		/note= "intracellular C-terminal domain"	
XX			
PN	W09704094-A1.		
XX			
PD	06-FEB-1997.		
XX			
PF	24-JUL-1996;	96MO-FR01167.	
XX			
PR	24-JUL-1995;	95FR-0008947.	
XX			
PA	(ADIR ) ADIR & CIE.		
XX			
PI	Jockers R, Marullo S, Strosberg AD;		
XX			
DR	WPI: 1997-132635/12.		
XX			
DR	N-PSDB; T79065-6.		
XX			

PT New nucleic acid encoding functional melatonin receptor of *Xenopus* -  
PT for screening for potential (ant)agonists useful for e.g. treating  
PT cardiovascular disease and cancer  
PS Claim 5; Page 30-31; 62pp; French.  
XX Sequences W25926-7 represent novel allelic *Xenopus laevis* melatonin  
CC receptor MEL-1a proteins. This sequence is a new form of the melatonin  
CC receptor designated MEL-1ab, also known as Mel 1-c(beta). The protein  
CC differs from previously known receptors by being 65 amino acids shorter  
CC and also differs throughout the sequence by 6 amino acids. The protein  
CC is encoded by 2 different alleles (179065-6) which differ in the 3',  
CC untranslated region (both sequences encode the same protein), which is  
CC thought to affect the half-life of the mRNA. The coding sequence was  
CC isolated from cDNA derived from *Xenopus* skin RNA and amplified using the  
CC primers 179067-76. The nucleotide sequence encodes a 7 transmembrane  
CC receptor involved in cellular signalling. MEL-1ab has been shown to  
CC modulate intracellular cGMP, esp. inhibiting its accumulation induced by  
CC an inhibitor of phosphodiesterase, but unlike the MEL-1a protein, it  
CC cannot inhibit adenylyl cyclase activity.  
XX  
SQ Sequence 354 AA;

Query Match 7.88; Score 118; DB 18; Length 354;  
Best Local Similarity 19.48; Pred. No. 1.2e-05;  
Matches 57; Conservative 64; Mismatches 119; Indels 54; Gaps 12;

OY 2 LESHLIIFLAVIOELGIFNGIIVVNGIDILKRRMAPL-DLLSCLAVSRIFQL 60  
DB 31 lsaalavalfivadv9-----nlvli---lsvlnkklqnaqlfivssiadlvav 83  
OY 61 FIFYVWVIFPIEFIRMSANCALLFTELMLWMLGVFYCAKVASVRHPLFIMLK 120  
DB 84 yypvllallafqgwclqnhcqlsgflmglsv-----lgsvfntalaincyichsl 138  
OY 121 RISKLV----PKMILGSLIVSMICVFHSKVGFW--PYFLRKFSQNAITQKEDTLAI 174  
DB 139 rydklfngfstfyigltwlcltialvpnfivgslyqdrflscdaq--lvssstcltv 196  
OY 175 QFESFAERSVPLLIPLFVALLIFSLGRTRQ-----MNTVA----- 213  
DB 197 vvvhlvplsvvtfcyl-rfwlvivqvkhrvqgdfkqrltptcdlnflmfvvlfaavc 255  
OY 214 -----GSRV--PGRGAP-ISALLSILPLILFYSHCMKIVPLSSLEKPIIR 256  
DB 256 waplntfiglavalnplhvapkipewlfvlsymayfnsclnaviygllnqnfik 309

RESULT 10  
R88409 standard; Protein: 420 AA.  
XX R88409;  
XX 15-AUG-1996 (first entry)  
XX High-affinity melatonin receptor.  
XX  
XX Melatonin receptor: G-protein-coupled receptor; glycosylation;  
KW disulfide bond; ligand binding pocket; phosphorylation;  
KW cytochrome c family heme binding site; melatonin receptor-agonist;  
KW melatonin receptor-antagonist; circadian rhythm disorder; jet-lag;  
KW day-night cycle disorder; ovulation; reproductive cycle; antibody;  
XX animal breeding; puberty; transgenic animal; drug screening.  
XX  
XX *Xenopus laevis*.  
XX  
XX Key Location/Qualifiers  
FH Domain 1..32  
FT Domain /note= "Extracellular domain"  
FT Modified-site 5..7  
FT /note= "N-glycosylation site"

FT Modified-site 16  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Domain 33..57  
FT /note= "Transmembrane region-I"  
FT Domain 57..68  
FT /note= "Intracellular loop"  
FT Region 67..72  
FT /note= "Conserved melatonin receptor motif"  
FT Domain 69..93  
FT /note= "Transmembrane region-II"  
FT Domain 94..107  
FT /note= "Extracellular loop"  
FT Disulfide-bond 105..182  
FT Domain 108..126  
FT /note= "Transmembrane region-III"  
FT Region 125..131  
FT /note= "Region used to construct primer"  
FT Domain 127..151  
FT /note= "Intracellular loop"  
FT Region 132..137  
FT /note= "Putative cytochrome-c family heme binding site"  
FT Modified-site 137  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Domain 152..171  
FT /note= "Transmembrane region-IV"  
FT Misc-difference 166  
FT /note= "Residue which may form ligand binding pocket"  
FT Domain 172..193  
FT /note= "Extracellular loop"  
FT Domain 194..220  
FT /note= "Transmembrane region-V"  
FT Misc-difference 204  
FT /note= "Residue which may form ligand binding pocket"  
FT Domain 221..243  
FT /note= "Intracellular loop"  
FT Domain 244..268  
FT /note= "Transmembrane region-VI"  
FT Region 252..259  
FT /note= "Region used to construct primer"  
FT Misc-difference 258  
FT /note= "Residue which may form ligand binding pocket"  
FT Domain 269..279  
FT /note= "Extracellular loop"  
FT Domain 280..301  
FT /note= "Transmembrane region-VII"  
FT Region 296..300  
FT /note= "Conserved melatonin receptor-1b motif"  
FT Domain 302..420  
FT /note= "Intracellular domain"  
FT Modified-site 320  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Modified-site 328  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Modified-site 362  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Modified-site 395  
FT /note= "Protein-kinase-C phosphorylation site"  
FT Modified-site 418  
FT /note= "Protein-kinase-C phosphorylation site"

XX W09535320-A1.  
XX PN 28-DEC-1995.  
XX PD 07-JUN-1995;  
XX PR 06-JUN-1995;  
XX PR 17-JUN-1994;  
XX PR 07-OCT-1994;  
XX (MASS-) MASSACHUSETTS GEN HOSPITAL.  
XX Reprint SM;  
PI

XX WPI: 1996-058368/06.  
DR N-PSDB: T09947.  
XX  
PT DNA encoding high affinity melatonin receptor one - used to identify  
PT receptor agonists or antagonists e.g. for regulating circadian  
PT rhythm disorders or reproductive cycles  
XX  
PS Claim 5; Fig 1; 115pp; English.

XX The sequence represents a high-affinity melatonin receptor (mol.wt.  
CC 47,424) from *Xenopus laevis*. The receptor is a membrane protein,  
CC coupled to guanine nucleotide binding proteins (G-proteins), and  
CC has 7 hydrophobic putative transmembrane domains. The N-terminus  
CC has an N-linked glycosylation site typical for G-protein-coupled  
CC receptors, and 2 Cys residues in the last 2 extracellular loops may  
CC form a stabilizing disulfide bond. pro residues in transmembrane  
CC domains IV, V and VI may introduce kinks in the alpha-helices to  
CC form of a ligand binding pocket. Phosphorylation sites in the  
CC C-tail may be involved in receptor regulation. Primers from the  
CC encoding DNA may be used for isolation of sheep, mouse and human  
CC receptor sequences. Receptor fragments which interact with  
CC melatonin, or specific antibodies, may be used as receptor-agonists  
CC or receptor-antagonists. Agonists such as jet-lag or day-night cycle  
CC circadian rhythm disorders as jet-lag or day-night cycle  
CC disorders, to control ovulation, or in alteration of reproductive  
CC cycles in seasonally breeding animals. Antagonists may be used to  
CC control the initiation or timing of puberty in humans. The  
CC receptor gene may also be expressed in a transgenic animal for use  
CC as a model system to screen agonists and antagonists.

XX Sequence 420 AA:

Query Match 7.8%; Score 118; DB 17; Length 420;  
Best Local Similarity 19.4%; Pred. No. 1.5e-05;  
Matches 57; Conservative 64; Mismatches 119; Indels 54; Gaps 12;

QY 2 LESHLITVFLAVIQPLIGFTNGIIVVNGIDILIRKRNAPL-DLLSCLAVSRIFQL 60  
DB 31 ltsalavvllfclvadvlg---nlvyl---lavlrrnkklqngnlfvvaladlvav 83  
QY 61 FIFVYVIVFEEFIEFICWNCALILFELNELMLATWIGVFCANVASRHHFLTMK 120  
DB 84 YPVPVILIAIFGNGWLTGNHCGISGFLMGV---IGSVNLTALAINRYCCHSL 138  
QY 121 RISKLV---PMWILGSLVSMICVFNHRYAGFMV---PYELRRKFSQNTIOKEDTLAI 174  
DB 139 rydklyngstwtcylgtlwtlltllatpntffvgslyqdpdlfacteq--tvesstlytic 196  
QY 175 QIFSFEVAFESVPILIFLFAVLLILFISLGRTRQ-----MRNTVA----- 213  
DB 197 vvvhflvplsvvfcgyl-rfwlvvlvqkhvtrgdffkqklctqldrlfmlfvfvlfavc 255  
QY 214 -----GSRV---PGRGAP-ISALLSILSLFLYFSHCWKIKVFLSLKFHRH 256  
DB 256 waplntflglavlnpflvapkfpewlfvlsayfmayfnsclnavlyvlnqnfik 309

RESULT 11  
W19105 standard; Protein; 310 AA.  
XX W19105;  
XX  
XX 26-AUG-1997 (first entry)  
XX  
XX Rat pheromone receptor VN4.  
DE  
XX Pheromone receptor; vomeronasal sensory neuron; social behaviour;  
KW maternal behaviour; reproductive behaviour; fertility;  
KW hormone secretion.  
XX

OS Rattus sp.  
XX  
FH Key Location/Qualifiers  
FT Domain 16..37  
FT /label= I  
FT /note= "Predicted position of transmembrane domain I"  
FT  
FT Domain 50..71  
FT /label= II  
FT /note= "Predicted position of transmembrane domain II"  
FT  
FT Domain 97..116  
FT /label= III  
FT /note= "Predicted position of transmembrane domain III"  
FT  
FT Domain 135..156  
FT /label= IV  
FT /note= "Predicted position of transmembrane domain IV"  
FT  
FT Domain 194..212  
FT /label= V  
FT /note= "Predicted position of transmembrane domain V"  
FT  
FT Domain 240..258  
FT /label= VI  
FT /note= "Predicted position of transmembrane domain VI"  
FT  
FT Domain 270..292  
FT /label= VII  
FT /note= "Predicted position of transmembrane domain VII"  
FT  
FT Misc-difference 261  
FT /note= "Ser-261 is additional to the amino acid residues deduced from the nucleotide sequence"  
FT  
FT Misc-difference 263..264  
FT /note= "the amino acid sequence deduced from the nucleotide sequence has a lysine residue between Met-263 and Phe-264"  
FT  
PN W09714790-A1.  
XX  
XX 24-APR-1997.  
PD  
XX  
XX 18-OCT-1996; 96MO-US16637.  
PF  
XX  
XX 19-OCT-1995; 95US-0005698.  
PR  
XX  
XX (UYCO ) UNITV COLUMBIA NEW YORK.  
PA  
XX  
XX Axel R, Dulac C;  
PI  
XX  
XX WPI: 1997-245107/22.  
DR N-PSDB: T69547.  
XX  
XX Nucleic acid molecule encoding vertebrate pheromone receptor -  
PT useful to identify modulators for control of reproductive and social  
PT behaviour, fertility and hormone secretion  
PS  
PS Disclosure: Page 79-80; 123pp; English.  
XX  
XX Rat pheromone receptors VN1, VN3, VN4, VN5, VN6, VN7 and VN2  
CC (W19103-09) are members of a novel family of presumed  
CC 7-transmembrane domain receptors that are evolutionarily independent  
CC of the odorant receptors of the main olfactory epithelium (MOE).  
CC Their amino acid sequences were deduced from cDNA clones (see also  
CC T69545-50. VN2 sequence not provided) obtd. by differential  
CC cloning, PCR and hybridisation from single vomeronasal sensory  
CC neurons (VSN). Individual neurons express different complements of  
CC the receptors. VN polypeptides can be expressed in host-vector  
CC systems for use in identifying modulators for control of maternal,  
CC reproductive and social behavior, to increase fertility, control  
CC hormone secretion and to regulate food uptake in humans and



CC	18-SEP-1997;	97US-0059263.
CC	18-SEP-1997;	97US-0059266.
PR	15-OCT-1997;	97US-0062125.
PR	17-OCT-1997;	97US-0062285.
PR	17-OCT-1997;	97US-0062287.
PR	21-OCT-1997;	97US-0063486.
PR	24-OCT-1997;	97US-0062814.
PR	24-OCT-1997;	97US-0062816.
PR	24-OCT-1997;	97US-0063045.
PR	24-OCT-1997;	97US-0063120.
PR	24-OCT-1997;	97US-0063121.
PR	24-OCT-1997;	97US-0063127.
PR	27-OCT-1997;	97US-0063128.
PR	27-OCT-1997;	97US-0063329.
PR	28-OCT-1997;	97US-0063327.
PR	28-OCT-1997;	97US-0063541.
PR	28-OCT-1997;	97US-0063542.
PR	28-OCT-1997;	97US-0063544.
PR	28-OCT-1997;	97US-0063549.
PR	28-OCT-1997;	97US-0063550.
PR	28-OCT-1997;	97US-0063564.
PR	29-OCT-1997;	97US-0063435.
PR	29-OCT-1997;	97US-0063704.
PR	29-OCT-1997;	97US-0063732.
PR	29-OCT-1997;	97US-0063738.
PR	29-OCT-1997;	97US-0063734.
PR	29-OCT-1997;	97US-0064215.
PR	29-OCT-1997;	97US-0064735.
PR	31-OCT-1997;	97US-0063870.
PR	31-OCT-1997;	97US-0064103.
PR	03-NOV-1997;	97US-0064248.
PR	07-NOV-1997;	97US-0064809.
PR	12-NOV-1997;	97US-0065186.
PR	17-NOV-1997;	97US-0065846.
PR	18-NOV-1997;	97US-0065693.
PR	21-NOV-1997;	97US-0066120.
PR	21-NOV-1997;	97US-0066364.
PR	24-NOV-1997;	97US-0066772.
PR	24-NOV-1997;	97US-0066466.
PR	24-NOV-1997;	97US-0066770.
PR	24-NOV-1997;	97US-0066511.
PR	24-NOV-1997;	97US-0066453.
PA	(GETH ) GENENTECH INC.	
XX		
XX		
PL	Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;	
XX	WPI: 1999-229533/19.	
DR	N-PSDB: X52222.	
XX		
PT	New isolated human genes and polypeptides used in, e.g. treatment of	
PT	gastrointestinal ulceration	
PS		
XX	Claim 12; Fig 19; 320pp; English.	
XX		
CC	Y13344-403 represent secreted and transmembrane human proteins.	
CC	The cDNA sequences are obtained from cDNA libraries, prepared from	
CC	fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.	
CC	The encoded polypeptides have specific uses based on their homology to	
CC	known polypeptides, e.g. PRO211 and PRO217 can be used for disorders	
CC	associated with the preservation and maintenance of gastrointestinal	
CC	mucosa and the repair of acute and chronic mucosal lesions	
CC	(e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal	
CC	ulceration and congenital microvillus atrophy), skin diseases associated	
CC	with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial	
CC	cancers such as lung squamous cell carcinoma of the vulva and gliomas),	
CC	potent effects on cell growth and development, diseases related to growth	
CC	or survival of nerve cells including Parkinson's disease, Alzheimer's	
CC	disease, ALS, neuropathies or cancer. PRO265 can be used as for	
CC	fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used	
CC	as a target for anti-tumor drugs. PRO533 may be used in the treatment	
CC	of Usher Syndrome or Atrophica areata. PRO269 can be used as an	
CC	anti-thrombotic agent; PRO287 polypeptides and portions may have	

[illegible]



FT		Modified-site	/label= "transmembrane-domain"
FT		670..677	
FT		/note= "tyrosine kinase phosphorylation site"	
XX	PN	MO200021996-A2.	
XX	PD	20-APR-2000.	
XX	PF	05-OCT-1999;	99WO-US23089.
XX	PR	13-OCT-1998;	98US-0104080.
XX	PA	(GETH ) GENENTECH INC.	
PI	Pi	Ashkenazi A, Goddard A, Gurney AL, Klein RD, Napier M, Wood WI;	
PI	Pi	Yuan J;	
DR	N-PSDB:	WPI: 2000-317943/27.	
DR	N-PSDB:	Z93701.	
PT		Composition for inhibiting neoplastic cell growth and treating cancers	
PT		of ovary, uterus, prostate, lung and bladder, comprises PRO211, PRO228,	
PT		PRO538, PRO172 or PRO182 polypeptide or their agonist	
PS		Claim 8; Figure 4; 122pp; English.	
CC		Compositions comprising a PRO211, PRO228, PRO538, PRO172 or PRO182	
CC		polypeptide or their agonists, mixed with a carrier is useful for	
CC		inhibiting neoplastic growth and treating tumors such as cancers of	
CC		breast, ovary, renal, colorectal, uterus, prostate, lung, bladder,	
CC		central nervous system, melanoma and leukemia.	
SQ		Sequence 690 AA:	
Query_Match	7.7%;	Score 117.5;	DB 21: Length 690;
Best Local Similarity	21.3%;	Pred. No. 3.3e-05;	
Matches 71;	Conservative 59;	Mismatches 114;	Indels 89; Gaps 16
OY	4 SHLIYELLAVIOFLGIFETNGIIVVANGIDLIKRRKMAPLDLLSCLAVSRIFQLFIF 63		
Db	404 nhlthfaImsspsjgldynllrtltqgl-----slclalc-----lftf 449		
OY	64 YVNVIYIEFIET-----IMCSANCAILLFINELEMLATWGVEFCAYASVRHP 113		
Db	450 w-----ffseIsgrtltnknlccsfllaelyvl---vgIntntnklfcslagllhy 499		
OY	114 LFTWLKRIRIKIPWM--ILGLLYVMICVFHSRYAGFM--VPFLRKFSQNATIONEDT 171		
Db	500 ffl-----aaIawmcleglnlyllvgyvlynk--gfInknfyIfgyLspavvgfsaa 550		
OY	172 LAIQ-----IFFSVAFESVPLLEFAVLILFFSLGRTRORNRNYA-- 213		
Db	551 lgyryygtctkcwstennflwfsgpacalllvnllaifgvilykvfirhtaglkevsdf 610		
OY	214 -GSNVPGRGAPISLLLSLTS----ELLTPSHCM---KYELSLKRIHRIFLPFILV 265		
Db	611 enIscarga--laIlflIgtwtlfyvlhvhasvtraylitvsnaql-gmfIlflfcvl 667		
OY	266 IGIVPSSHLLILGNPKLKONAKFKLLHSKCC 298		
Db	668 -----srklgeeyrlfknpcc 685		
RESULT 15			
P90554			
ID	P90554 standard; protein; 349 AA.		
AC	P90554:		
XX			
DT	16-FEB-1990 (first entry)		
DE	Bovine rhodopsin.		

```

XX  Bovine rhodopsin; 5HT1c receptor.
KW
XX  Bovidae.
OS
XX  Key
XX  Location/Qualifiers
XX  38..62
XX  /label=I
XX  /note="transmembrane domain."
XX  74..94
XX  /label=II
XX  /note="transmembrane domain."
XX  111..133
XX  /label=III
XX  /note="transmembrane domain."
XX  153..174
XX  /label=IV
XX  /note="transmembrane domain."
XX  201..224
XX  /label=V
XX  /note="transmembrane domain."
XX  253..276
XX  /label=VI
XX  /note="transmembrane domain."
XX  290..310
XX  /label=VII
XX  /note="transmembrane domain."
XX  236
XX  /note="cytoplasmic loop."
XX
XX  W08918149-A.
XX  08-SEP-1989.
XX
XX  28-FEB-1989; 89MO-US00808.
XX
XX  29-FEB-1988; 88US-0162654.
XX
XX  (UYCO-) COLUMBIA UNIVERSITY.
XX
XX  Axel R, Jessell TM;
XX
XX  WPI; 1989-278308/38.
XX
XX  DNA encoding serotonin 5HT1c receptor - used for producing protein,
XX  antibodies and probes for studying receptor binding and screening drugs.
XX
XX  Disclosure: Fig 11; 84pp; English.
XX
XX  Sequence codes for bovine rhodopsin.
XX  See also P90549-P90554, P92111 and N90955.
XX
XX  Sequence 349 AA;

Query Match 7.6%; Score 116; DB 10; Length 349;
Best Local Similarity 19.9%; Pred. No. 2, 1e-05;
Matches 63; Conservative 66; Mismatches 133; Indels 54; Gaps 15;

QY 9 YELLAVIOFL---LGITNGIIVVNGIDLIKHKM-APDLLSCLAVSRIFLOLFIY 64
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 37 fmlaaymflimlqiflflitlyc---vqhkikrplnylllnlavaadlfmvgft 92
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 65 VNIVIVFIEFIWCSANCAILFT---NELIWM---LATWLGVPFYCAKVASVR---H 112
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 93 tclvtalhyfvtgptgcmlgegfaclgelaivlaieryyvvckpmnsffgenh 152
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 PL-----FWLKMRIKLVPMMLGSLVY--SMICVHSHKYGAFWVPLRKRFSONATI 166
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 153 alngvafv--malacaap--plvgwarylpegmqc---scgldytpb----- 195
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 OKEDTIAIOIFSVAFSPFLILFVALLIFSLGRHTROMKNTVAGSRVPGCAPISA 226
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 196 eetneafviymfvvhfildllylffcygqlvfvkeaaaxqggesatqkaekvtrmv 255
QY 227 LSLISFLILYFESHCMIKVFLSSLSKFIHRRFIPL---FPIVIGIYPSCHSILILGNP 282
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 256 lhmviaflicwlpyaqvaaylftthgsdfgplfmltpatfaktcsavy---npviyimmnk 312
QY 283 KKKONAKKFLILHSKCC 298
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 313 qf---fncmvtlicc 324

```

Search completed: March 15, 2001, 12:58:48  
Job time: 1569 sec